

**PCT**ORGANISATION MONDIALE DE LA PROPRIÉTÉ INTELLECTUELLE  
Bureau international

## DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITE DE COOPERATION EN MATIERE DE BREVETS (PCT)

(51) Classification internationale des brevets <sup>6</sup> : <b>C12Q 1/68</b>	<b>A1</b>	(11) Numéro de publication internationale: <b>WO 99/46403</b>  (43) Date de publication internationale: 16 septembre 1999 (16.09.99)
(21) Numéro de la demande internationale: <b>PCT/FR99/00547</b>  (22) Date de dépôt international: <b>11 mars 1999 (11.03.99)</b>  (30) Données relatives à la priorité: <b>98/02997</b> <b>11 mars 1998 (11.03.98)</b> <b>FR</b>  (63) Apparentée par "continuation" (CON) ou par "continuation-in-part" (CIP) à une demande antérieure <b>US</b> <b>09/046,920 (CIP)</b> <b>Déposée le</b> <b>24 mars 1998 (24.03.98)</b>  (71) Déposant (pour tous les Etats désignés sauf US): <b>EXON-</b> <b>HIT THERAPEUTICS S.A. [FR/FR]; 63-65, boulevard</b> <b>Masséna, F-75013 Paris (FR).</b>  (72) Inventeurs; et (75) Inventeurs/Déposants (US seulement): <b>SCHWEIGHOFFER,</b> <b>Fabien [FR/FR]; 38, avenue Paul Déroulède, F-94300</b> <b>Vincennes (FR). BRACCO, Laurent [FR/FR]; 10, rue du</b> <b>Moulin des Prés, F-75013 Paris (FR). TOCQUE, Bruno</b> <b>[FR/FR]; 58, boulevard Saint-Denis, F-92400 Courbevoie</b> <b>(FR).</b>	(74) Mandataires: <b>GUTMANN, Ernest etc.; Ernest Gutmann-Yves</b> <b>Plasseraud S.A., 3, rue Chauveau-Lagarde, F-75008 Paris</b> <b>(FR).</b>  (81) Etats désignés: <b>AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR,</b> <b>BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD,</b> <b>GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP,</b> <b>KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK,</b> <b>MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI,</b> <b>SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW,</b> <b>brevet ARIPO (GH, GM, KE, LS, MW, SD, SL, SZ, UG,</b> <b>ZW), brevet eurasién (AM, AZ, BY, KG, KZ, MD, RU, TJ,</b> <b>TM), brevet européen (AT, BE, CH, CY, DE, DK, ES, FI,</b> <b>FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), brevet OAPI</b> <b>(BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE,</b> <b>SN, TD, TG).</b>  <b>Publiée</b> <i>Avec rapport de recherche internationale.</i> <i>Avant l'expiration du délai prévu pour la modification des</i> <i>revendications, sera republiée si des modifications sont</i> <i>reçues.</i>	
<p>(54) Title: <b>QUALITATIVE DIFFERENTIAL SCREENING</b></p> <p>(54) Titre: <b>CRIBLAGE DIFFERENTIEL QUALITATIF</b></p> <p>(57) Abstract</p> <p>The invention concerns a method for identifying and/or cloning nucleic acid regions representing qualitative differences associated with alternative splicing events and/or with insertions, deletions located in RNA transcribed genome regions, between two physiological situations, comprising either hybridisation of RNA derived from the test situation with cDNA's derived from the reference situation and/or reciprocally, or double-strand hybridisation of cDNA derived from the test situation with cDNA's derived from the reference situation; and identifying and/or cloning nucleic acids representing qualitative differences. The invention also concerns compositions or banks of nucleic acids representing qualitative differences between two physiological situations, obtainable by the above method, and their use as probe, for identifying genes or molecules of interest, or still for example in methods of pharmacogenomics, and profiling of molecules relative to their therapeutic and/or toxic effects. The invention further concerns the use of dysregulation of splicing RNA as markers for predicting molecule toxicity and/or efficacy, and as markers in pharmacogenomics.</p> <p>(57) Abrégé</p> <p>L'invention concerne un procédé d'identification et/ou de clonage de régions d'acides nucléiques représentatives de différences qualitatives associées à des événements d'épissages alternatifs et/ou à des insertions, délétions se trouvant dans des régions du génome transcrites en ARN, entre deux situations physiologiques, comprenant soit l'hybridation d'ARN provenant de la situation test avec les ADNc provenant de la situation de référence et/ou réciproquement, soit l'hybridation d'ADNc double-brin provenant de la situation test avec les ADNc provenant de la situation de référence et l'identification et/ou le clonage d'acides nucléiques représentatifs des différences qualitatives. L'invention concerne également des compositions ou banques d'acides nucléiques représentatifs de différences qualitatives entre deux situations physiologiques, susceptibles d'être obtenues par le procédé décrit ci-dessus, ainsi que leur utilisation comme sonde, pour l'identification de gènes ou molécules d'intérêt, ou encore par exemple dans des méthodes de pharmacogénomique, et de profilage de molécules vis à vis de leurs effets thérapeutiques et/ou toxiques. L'invention concerne aussi l'utilisation des dysrégulations de l'épissage des ARN comme marqueurs de prédiction de la toxicité et/ou de l'efficacité de molécules, ainsi que comme marqueurs de pharmacogénomique.</p>		

# **UNIQUEMENT A TITRE D'INFORMATION**

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publiant des demandes internationales en vertu du PCT.

AL	Albanie	ES	Espagne	LS	Lesotho	SI	Slovénie
AM	Arménie	FI	Finlande	LT	Lituanie	SK	Slovaquie
AT	Autriche	FR	France	LU	Luxembourg	SN	Sénégal
AU	Australie	GA	Gabon	LV	Lettonie	SZ	Swaziland
AZ	Azerbaïdjan	GB	Royaume-Uni	MC	Monaco	TD	Tchad
BA	Bosnie-Herzégovine	GE	Géorgie	MD	République de Moldova	TG	Togo
BB	Barbade	GH	Ghana	MG	Madagascar	TJ	Tadjikistan
BE	Belgique	GN	Guinée	MK	Ex-République yougoslave de Macédoine	TM	Turkménistan
BF	Burkina Faso	GR	Grèce	ML	Mali	TR	Turquie
BG	Bulgarie	HU	Hongrie	MN	Mongolie	TT	Trinité-et-Tobago
BJ	Bénin	IE	Irlande	MR	Mauritanie	UA	Ukraine
BR	Brsél	IL	Israël	MW	Malawi	UG	Ouganda
BY	Bélarus	IS	Islande	MX	Mexique	US	Etats-Unis d'Amérique
CA	Canada	IT	Italie	NE	Niger	UZ	Ouzbékistan
CF	République centrafricaine	JP	Japon	NL	Pays-Bas	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norvège	YU	Yougoslavie
CH	Suisse	KG	Kirghizistan	NZ	Nouvelle-Zélande	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	République populaire démocratique de Corée	PL	Pologne		
CM	Cameroun	KR	République de Corée	PT	Portugal		
CN	Chine	KZ	Kazakstan	RO	Roumanie		
CU	Cuba	LC	Sainte-Lucie	RU	Fédération de Russie		
CZ	République tchèque	LI	Liechtenstein	SD	Soudan		
DE	Allemagne	LK	Sri Lanka	SE	Suède		
DK	Danemark	LR	Libéria	SG	Singapour		
EE	Estonie						

### CRIBLAGE DIFFÉRENTIEL QUALITATIF

Cette invention se rapporte aux domaines techniques de la biotechnologie, de la médecine, de la biologie et de la biochimie. Ses applications concernent les domaines de la santé humaine, animale et végétale. Plus particulièrement, l'invention permet d'identifier des séquences d'acides nucléiques permettant de concevoir de nouveaux cribles pour molécules d'intérêt thérapeutique, de nouveaux outils de thérapie génique ainsi que d'apporter des indications sur le potentiel toxique et le suivi d'efficacité de molécules et des informations de pharmacogénomique.

La présente invention décrit notamment une série de techniques originales d'identification de séquences d'acides nucléiques basée sur la mise en évidence des différences qualitatives entre les ARN issus de deux contextes différents que l'on désire comparer, en particulier issus d'un tissu ou d'un organe malade et leur équivalent sain. Plus précisément, ces techniques sont destinées à cloner spécifiquement les introns et les exons alternatifs épissés différemment entre une situation pathologique et un état sain ou entre deux situations physiologiques que l'on désire comparer. Ces différences qualitatives au sein des ARNs peuvent également provenir d'altération(s) du génome, de type insertions ou délétions dans des régions qui seront transcrites en ARN. Cette série de techniques est identifiée par l'acronyme DATAS : Differential Analysis of Transcripts with Alternative Splicing.

La caractérisation des altérations de l'expression génétique qui président ou sont associées à une pathologie donnée suscite un espoir important de découvrir de nouvelles cibles thérapeutiques et de nouveaux outils diagnostiques. Toutefois, l'identification d'une séquence d'ADN génomique ou complémentaire, qu'elle ait lieu par clonage positionnel ou par des techniques de criblage différentiel quantitatif, n'apporte que peu ou pas d'information sur la fonction et encore moins sur les domaines fonctionnels mis en jeu dans les dérégulations liées à la pathologie étudiée. La présente invention décrit une série de techniques originales qui visent à identifier les différences d'épissages des ARNs qui existent entre deux situations

physiopathologiques distinctes. L'identification de ces différences apporte des informations sur les différences qualitatives et non sur les différences quantitatives, comme c'est le cas pour les techniques décrites jusqu'à présent. L'ensemble des techniques présentées dans la présente invention  
5 sont donc regroupées sous l'appellation "criblage différentiel qualitatif" ou DATAS. Les méthodes de l'invention sont utilisables pour l'identification de nouvelles cibles ou produits thérapeutiques, pour la préparation d'outils de recherche génétique et/ou d'outils de diagnostic, pour la construction de banques d'acides nucléiques, et dans des méthodes de détermination du  
10 profil toxicologique ou de l'efficacité d'un composé par exemple.

Un premier objet de l'invention réside plus particulièrement dans un procédé d'identification et/ou de clonage de régions d'acides nucléiques représentatives de différences génétiques qualitatives entre deux échantillons biologiques, comprenant une étape d'hybridation d'une  
15 population d'ARN ou d'ADNc double-brin provenant d'un premier échantillon biologique avec une population d'ADNc provenant d'un deuxième échantillon biologique (Figure 1A).

Comme indiqué ci-avant, les différences génétiques qualitatives peuvent être dues à des modifications d'épissage des ARN ou à des  
20 délétions et/ou insertions dans des régions du génome qui sont transcrites en ARN.

Dans un premier mode de réalisation, il s'agit d'une hybridation entre une population d'ARN provenant d'un premier échantillon biologique et une population d'ADNc (simple-brin ou double-brin) provenant d'un deuxième  
25 échantillon biologique.

Dans un autre mode de réalisation, il s'agit d'une hybridation entre une population d'ADNc double-brin provenant d'un premier échantillon biologique et une population d'ADNc (double-brin ou, préférentiellement simple-brin) provenant d'un deuxième échantillon biologique.

30 Un objet plus particulier de l'invention réside dans un procédé d'identification de régions d'acides nucléiques épissées différemment

entre deux situations physiologiques, comprenant l'hybridation d'une population d'ARN ou d'ADNc double-brin provenant d'une situation test avec une population d'ADNc provenant d'une situation de référence et l'identification d'acides nucléiques correspondant à des épissages  
5 différentiels.

Un autre objet de l'invention concerne un procédé de clonage d'acides nucléiques épissés différemment entre deux situations physiologiques, comprenant l'hybridation d'une population d'ARN ou d'ADNc double-brin provenant de la situation test avec une population d'ADNc  
10 provenant de la situation de référence et le clonage d'acides nucléiques correspondant à des épissages différentiels.

Dans un mode particulier de mise en œuvre, le procédé d'identification et/ou de clonage d'acides nucléiques de l'invention comprend deux hybridations parallèles :

15 (a) l'hybridation des ARN provenant du premier échantillon (situation test) avec les ADNc provenant du deuxième échantillon (situation de référence);

(b) l'hybridation des ARN provenant du deuxième échantillon (situation de référence) avec les ADNc provenant du premier échantillon  
20 (situation test); et

(c) l'identification et/ou le clonage, à partir des hybrides formés en (a) et (b) d'acides nucléiques correspondant à des différences génétiques qualitatives.

La présente invention concerne également la préparation de  
25 banques d'acides nucléiques, les acides nucléiques et banques ainsi obtenus, ainsi que les utilisations de ces matériels dans tous les domaines de la biologie/biotechnologie, comme illustré plus loin.

A cet égard, l'invention est également relative à un procédé de préparation de compositions ou banques d'acides nucléiques profilées,  
30 représentatives des différences qualitatives existant entre deux échantillons biologiques, comprenant une étape d'hybridation d'une population d'ARN

provenant d'un premier échantillon biologique avec une population d'ADNc  
provenant d'un deuxième échantillon biologique.

L'invention concerne en outre une méthode de profilage  
("profiling") d'une composition d'ADNc, comprenant une étape d'hybridation  
5 de cette composition avec une population d'ARN, ou inversement.

Comme indiqué ci avant, la présente invention concerne en  
particulier des méthodes d'identification et de clonage d'acides nucléiques  
représentatifs d'un état physiologique. En outre, les acides nucléiques  
identifiés et/ou clonés représentent les qualités d'un état physiologique en ce  
10 sens que ces acides nucléiques sont généralement en grande partie  
impliqués dans l'état physiologique observé. De ce fait, les méthodes  
qualitatives de l'invention donnent accès directement aux éléments  
génétiques ou à leur produit protéique, ayant un rôle fonctionnel dans le  
développement d'un état physiopathologique.

15 Les méthodes selon l'invention reposent en partie sur une  
étape originale d'hybridation croisée entre des ARN et des ADNc de  
situations physiologiques différentes. Cette ou ces hybridations croisées  
permettent avantageusement de mettre en évidence, dans les hybrides  
formés, des régions non appariées, c'est-à-dire des régions présentes dans  
20 les ARNs dans une situation physiologique donnée et pas dans les ARNs  
dans une autre situation physiologique. Ces régions correspondent  
essentiellement à des épissages alternatifs caractéristiques d'un état  
physiologique, mais peuvent également refléter des altérations génétiques  
de type insertions ou délétions, et constituent ainsi des éléments génétiques  
25 particulièrement utiles sur le plan thérapeutique ou diagnostic comme  
expliqué ci-après. L'invention consiste donc notamment à conserver les  
complexes formés après hybridation(s) croisée(s), afin d'en extraire les  
régions correspondant à des différences qualitatives. Cette méthodologie se  
différencie des techniques de soustractions quantitatives connues de  
30 l'homme de l'art ( Sargent and Dawid (1983), Science, 222, 135-139 ; Davis  
et.al. (1984), PNAS, 81, 2194-2198 ; Duguid and Dinauer (1990) Nuc. Acid

Res., 18, 2789-2792 ; Diatchenko et.al. (1996) PNAS, 93, 6025-6030), qui après hybridation(s), éliminent les hybrides formés pour ne conserver que les acides nucléiques non complexés.

L'invention concerne donc en premier lieu un procédé  
5 d'identification d'acides nucléiques d'intérêt comprenant l'hybridation entre les ARN d'un échantillon test et les ADNc d'un échantillon de référence. Cette hybridation permet de mettre en évidence, au sein des complexes formés, des différences génétiques qualitatives entre les situations testées, et ainsi d'identifier et/ou de cloner par exemple les épissages  
10 caractéristiques de la situation test.

Selon une première variante de l'invention, le procédé permet donc de générer une population d'acides nucléiques caractéristiques des épissages de l'état physiologique test par rapport à l'état de référence (Figure 1A, 1B). Comme indiqué ci-après, cette population peut être utilisée pour le  
15 clonage et la caractérisation des acides nucléiques, leur utilisation en diagnostic, criblage, thérapeutique ou pour la production d'anticorps ou de fragments protéiques ou de protéines entières. Cette population peut également servir à la constitution de banques utilisables dans différents domaines d'applications illustrés plus loin et à la réalisation de sondes  
20 marquées (Figure 1D).

Selon une autre variante de l'invention, le procédé comprend une première hybridation telle que décrite ci avant et une seconde hybridation, en parallèle, entre les ARN provenant de la situation de référence et les ADNc provenant de la situation test. Cette variante est  
25 particulièrement avantageuse puisqu'elle permet de générer deux populations d'acides nucléiques, l'une représentant les qualités de la situation test par rapport à la situation de référence, et l'autre les qualités de la situation de référence par rapport à la situation test (Figure 1C). Ces deux populations peuvent également être utilisées comme source d'acides  
30 nucléiques, ainsi que comme banques témoignant de l'empreinte génétique d'une situation physiologique donnée, comme détaillé plus loin (Figure 1D).

La présente invention peut être appliquée à tous types d'échantillons biologiques. En particulier, l'échantillon biologique peut être toute cellule, organe, tissu, prélèvement, biopsie, etc., contenant des acides nucléiques. S'agissant d'un organe, tissu ou biopsie, ils sont éventuellement mis en culture de manière à permettre l'accès aux cellules qui les composent. Il peut s'agir d'échantillons provenant de mammifères (en particulier l'homme), de végétaux, de bactéries ou de cellules eucaryotes inférieures (levures, cellules fongiques, etc.). Des exemples de matériels sont en particulier une biopsie de tumeur, une biopsie de plaques neurodégénératives ou d'aires cérébrales présentant des atteintes neurodégénératives, un échantillon de peau, un échantillon de cellules sanguines obtenues après prise de sang, une biopsie colorectale, des biopsies issues de lavages pulmonaires, etc. Des exemples de cellules sont notamment les cellules musculaires, hépatiques, fibroblastes, nerveuses, de l'épiderme, du derme, des cellules sanguines comme les lymphocytes B, T, les mastocytes, les monocytes, les granulocytes, les macrophages.

Comme indiqué ci avant, le criblage différentiel qualitatif selon la présente invention permet d'identifier des acides nucléiques caractéristiques d'une situation physiologique donnée (situation B) par rapport à une situation physiologique de référence (situation A), en vue de leur clonage ou autres utilisations. A titre illustratif, les situations physiologiques A et B étudiées peuvent être les suivantes :

SITUATION A	SITUATION B
échantillon sain	échantillon pathologique
échantillon sain	échantillon apoptotique
échantillon sain	échantillon après infection virale
échantillon sensible à X	échantillon résistant à X



échantillon non traité	échantillon traité (par exemple par composé toxique)
échantillon non différencié	échantillon ayant subi une différenciation cellulaire ou tissulaire

### Populations d'ARN

Pour la mise en œuvre de la présente invention, il est possible d'utiliser des ARN totaux ou les ARN messagers. Ces ARN peuvent être préparés par toutes méthodes classiques de biologie moléculaire, bien connues de l'homme du métier. Ces méthodes comprennent généralement une lyse des cellules ou tissu ou échantillons et l'isolement des ARNs par des techniques d'extraction. Il peut s'agir en particulier d'un traitement au moyen d'agents chaotropiques tels que le thiocyanate de guanidium (qui détruit les cellules et protège les ARN) suivi d'une extraction des ARN au moyen de solvants (phénol, chloroforme par exemple). De telles méthodes sont bien connues de l'homme du métier (voir Maniatis et al., Chomczynski et al., Anal. Biochem. 162 (1987) 156). Ces méthodes peuvent être aisément pratiquées en utilisant des kits disponibles dans le commerce tels que par exemple le kit US73750 (Amersham) ou le kit Rneasy (Quiagen) pour les ARN totaux. Il n'est pas nécessaire que les ARN utilisés soient parfaitement purs, et notamment il n'est pas gênant que des traces d'ADN génomique ou d'autres composants cellulaires (protéine, etc.) subsistent dans les préparations, dès lors qu'ils n'affectent pas significativement la stabilité des ARNs et que les modes de préparation entre les différents échantillons à comparer soient les mêmes. En outre, de manière facultative, il est possible d'utiliser non pas des préparations d'ARN totaux mais des préparations d'ARN messagers. Ceux-ci peuvent être isolés, soit directement à partir de l'échantillon biologique soit à partir des ARN totaux, au moyen de séquences polyT, selon les méthodes classiques. L'obtention d'ARN messagers peut à

cet égard être réalisée au moyen de kits commerciaux tels que par exemple le kit US72700 (Amersham) ou le kit utilisant des billes oligo-(dT) (Dyna). Un mode avantageux de préparation d'ARN consiste à extraire les ARN cytosoliques puis les ARN polyA<sup>+</sup> cytosoliques. Des kits permettant la  
5 préparation sélective d'ARN cytosoliques non contaminés avec des ARN pré-messagers porteurs d'exons et d'introns non épissés sont disponibles dans le commerce. C'est le cas notamment des kits Rneasy commercialisés par Qiagen (exemple de référence : 74103). Les ARN peuvent également être obtenus directement à partir de banques ou autres échantillons préparés  
10 à l'avance et/ou accessibles dans des collections, conservés dans des conditions appropriées.

Généralement, les préparations d'ARN utilisées comprennent avantageusement au moins 0,1 µg d'ARN, de préférence au moins 0,5 µg d'ARN. Les quantités peuvent varier selon les cellules et les méthodes  
15 utilisées, sans modifier la mise en oeuvre de la présente invention. Pour obtenir des quantités d'ARN suffisantes (de préférence 0,1 µg au moins), il est généralement recommandé d'utiliser un échantillon biologique comprenant au moins 10<sup>5</sup> cellules. A cet égard, une biopsie classique comprend généralement entre 10<sup>5</sup> et 10<sup>8</sup> cellules, et une culture cellulaire sur  
20 boîte de pétri classique (diamètre 6-10 cm) comporte de l'ordre de 10<sup>6</sup> cellules, ce qui permet d'obtenir aisément des quantités d'ARN suffisantes.

Les préparations d'ARN peuvent être utilisées extemporanément ou être conservées, de préférence au froid, en solution ou congelées, pour des utilisations ultérieures.

25

#### Populations d'ADNc

Les ADNc utilisés dans le cadre de la présente invention peuvent être obtenus par transcription inverse selon les techniques classiques de biologie moléculaire. On peut se référer notamment à Maniatis  
30 et al. La transcription inverse est généralement réalisée en utilisant une enzyme, transcriptase inverse ("reverse transcriptase") et une amorce.

A cet égard, de nombreuses transcriptases inverses ont été décrites dans la littérature et sont disponibles dans le commerce (Kit 1483188, Boehringer). On peut citer à titre d'exemples les transcriptases inverses les plus souvent utilisées comme celles du virus aviaire AMV (Avian Myeloblastosis Virus) et du virus de leucémie murine MMLV (Moloney Murine Leukemia Virus). Il convient également de citer certaines DNA polymérase thermostables douées d'activité transcriptase inverse telles celles isolées de *Thermus flavus* et de *Thermus thermophilus* HB-8 (commerciallement disponibles; références Promega M1941 et M2101). Selon une variante  
10 avantageuse, on utilise pour la mise en oeuvre de la présente invention la transcriptase inverse d'AMV puisque cette enzyme, fonctionnant à 42°C (contrairement à celle de MMLV qui fonctionne à 37°C), déstabilise certaines structures secondaires des ARN qui pourraient bloquer l'élongation, et permet ainsi la transcription inverse d'ARN de longueur importante, et permet  
15 d'avoir des préparations d'ADNc représentant les ARN avec une grande fidélité et une grande efficacité.

Selon une autre variante avantageuse de l'invention, on utilise une transcriptase inverse dépourvue d'activité RNaseH. L'utilisation de ce type d'enzyme offre plusieurs avantages, et notamment celui d'augmenter le  
20 rendement de synthèse des ADNc et de prévenir toute dégradation des ARN, qui seront ensuite engagés dans des hétéroduplex avec les ADNc néosynthétisés, permettant ainsi éventuellement de se passer de l'extraction phénolique de ceux ci. Les transcriptases inverses dépourvues d'activité RNaseH peuvent être préparées à partir de toute transcriptase inverse par  
25 délétion(s) et/ou mutagénèse. En outre, de telles enzymes sont également disponibles dans le commerce (par exemple Life Technologies, référence 18053-017).

Les conditions de mise en oeuvre des transcriptases inverses (concentration et température) sont bien connues de l'homme du métier. En  
30 particulier, on utilise généralement de 10 à 30 Unités d'enzyme par réaction, en présence d'une concentration optimale en  $Mg^{2+}$  de 10 mM.

La ou les amorces utilisées pour la transcription inverse peuvent être de natures différentes. Il peut s'agir en particulier d'un oligonucléotide aléatoire ("random") comprenant de préférence entre 4 et 10 nucléotides, avantageusement un hexanucléotide. L'utilisation de ce type d'amorce aléatoire a été décrite dans la littérature et permet d'initier la transcription inverse en différentes positions au hasard au sein des molécules d'ARN. Cette technique est surtout employée pour la transcription inverse d'ARN totaux (c'est-à-dire comprenant les ARNm, les ARNt et les ARNr notamment). Dans le cas où l'on souhaite réaliser la transcription inverse des ARNm seulement, il est avantageux d'utiliser comme amorce un oligonucléotide oligo dT, qui permet d'initier la transcription inverse à partir des queues polyA spécifiques des ARN messagers. L'oligonucléotide oligo dT peut comprendre de 4 à 20-mères, avantageusement de 15-mères environ. L'emploi de ce type d'amorce constitue un mode de réalisation préféré de l'invention. D'autre part, il peut être avantageux d'utiliser pour la transcription inverse une amorce marquée. Ceci peut en effet permettre de reconnaître et/ou de sélectionner et/ou de trier ultérieurement les ARN des ADNc. Ceci peut également permettre d'isoler les hétéroduplex ARN/ADN dont la formation représente une étape clef de l'invention. Le marquage de l'amorce peut consister en tout système de type ligand-récepteur, c'est-à-dire permettant par affinité de séparer les molécules portant l'amorce. Il peut s'agir par exemple d'un marquage par la biotine, qui peut être séparé par tout support (bille, colonne, plaques, etc.) sur lequel est fixée la streptavidine. Tout autre système de marquage permettant cette séparation sans affecter les propriétés d'amorce peut être utilisé de manière équivalente.

Dans les conditions habituelles de mise en oeuvre, cette transcription inverse génère des ADN complémentaires (ADNc) simple-brins. Ceci constitue un premier mode avantageux de la présente invention.

Dans une deuxième variante de mise en oeuvre, la transcription inverse est réalisée de manière à préparer des ADNc double-brins. Pour ce

faire, après transcription du premier brin d'ADNc, le deuxième brin peut être généré selon les techniques classiques de biologie moléculaire faisant intervenir des enzymes de modification de l'ADN comme l'ADN Ligase du phage T4, l'ADN polymérase I et l'ADN polymérase du phage T4.

5 Les préparations d'ADNc peuvent être utilisées extemporanément ou être conservées, de préférence au froid, en solution ou congelées, pour des utilisations ultérieures.

#### Hybridations

10 Comme expliqué ci-avant, les méthodes selon l'invention reposent en partie sur une étape originale d'hybridation croisée entre les ARN et les ADNc provenant d'échantillons biologiques dans des situations physiologiques ou d'origines différentes. Dans un mode de réalisation préféré, l'hybridation selon l'invention est avantageusement réalisée en  
15 phase liquide. En outre, elle peut être effectuée dans tout dispositif approprié, tel que par exemple des tubes (Eppendorff, par exemple), des plaques, ou tout autre support adapté et couramment utilisé en Biologie Moléculaire. L'hybridation est avantageusement réalisée dans des volumes compris entre 10 et 1000 µl, par exemple entre 10 et 500 µl. Il est entendu  
20 que le dispositif utilisé et les volumes utilisés peuvent être aisément adaptés par l'homme du métier. Les quantités d'acides nucléiques utilisées pour l'hybridation sont également connues de l'homme du métier. En général, il est suffisant d'utiliser des microgrammes d'acides nucléiques, par exemple de l'ordre de 0,1 à 100 µg.

25 Un élément plus important dans la mise en oeuvre des hybridations réside dans les quantités respectives d'acides nucléiques utilisées. Ainsi, il est possible d'utiliser les acides nucléiques dans un rapport ADNc/ARN variant de 50 à 0,02 environ, de préférence de 40 à 0,1. De manière plus particulièrement avantageuse, on préfère que le rapport  
30 ADNc/ARN soit proche ou supérieur à 1. En effet, dans ces expériences, l'ARN constitue le composé test ("tester") et l'ADNc constitue le porteur

("driver"). De ce fait, dans le but d'améliorer la spécificité de la méthode, il est préférable d'opérer dans des conditions où le "driver" se trouve en excès par rapport au "tester". En effet, dans ces conditions, l'effet de coopérativité entre les acides nucléiques joue et les appariements non-parfaits sont fortement défavorisés. De ce fait, les seuls mésappariements qui apparaissent sont  
5 généralement dus à la présence de régions dans les ARN "tester" qui n'existent pas dans l'ADNc "driver" et qui sont donc spécifiques. Pour favoriser la spécificité du procédé, l'hybridation est donc avantageusement réalisée à un rapport ADNc/ARN compris entre 1 environ et 10 environ. Il est  
10 bien entendu que ce rapport peut être adapté par l'homme du métier selon les conditions du procédé (quantités d'acides nucléiques disponibles, situations physiologiques, but poursuivi, etc.). Les autres paramètres de l'hybridation (temps, température, force ionique) sont également adaptables par l'homme du métier. De manière générale après dénaturation des "tester" et "driver" (par chauffage par exemple), l'hybridation est réalisée pendant  
15 environ 2 à 24 heures, à une température de 37°C environ (éventuellement soumise à des sauts de température comme décrit plus loin), et dans des conditions standard de force ionique (pouvant varier de 0,1 à 5M NaCl par exemple). Il est connu que la force ionique est un des facteurs déterminant la  
20 stringence d'une hybridation, notamment dans le cas d'hybridation sur support solide.

Selon un mode de mise en oeuvre particulier de l'invention, l'hybridation est réalisée en émulsion phénolique, par exemple selon la technique PERT ("Phenol Emulsion DNA Reassociation Technique) décrite  
25 par Kohne D.E. et al. (Biochemistry, Vol. 16, N° 24, pp 5329-5341, 1977). Avantageusement, on utilise dans le cadre de la présente invention l'hybridation en émulsion phénolique maintenue par thermocycles (sauts de température de 37°C environ à 60/65°C environ) et non par agitation, selon la technique décrite par Miller et Riblet (NAR 23 (1995) 2339). Toute autre  
30 technique d'hybridation en phase liquide, notamment en émulsion, peut être utilisée dans le cadre de la présente invention. Ainsi, dans un autre mode

particulièrement avantageux, l'hybridation est réalisée dans une solution contenant 80% de formamide, à une température de 40°C par exemple.

L'hybridation peut également se faire avec l'un des partenaires immobilisé sur un support. Avantageusement, c'est l'ADNc qui est immobilisé. Cela peut être réalisé en tirant profit du marquage dont peuvent faire l'objet les ADNc (voir ci-dessus) notamment grâce à des amorces biotinylées. Les groupements biotine sont mis en présence de billes magnétiques sur lesquelles sont fixées des molécules de streptavidine. Les ADNc peuvent ensuite être maintenus grâce à un aimant au contact d'un filtre ou d'un puits de plaque de microtitration. Les ARN sont ensuite, dans les conditions de force ionique requise, mis en présence des ADNc. Les ARN non appariés sont éliminés par lavage. Les ARN hybridés ainsi que les ADNc sont récupérés par retrait du champ magnétique.

Dans le cas où l'ADNc est double-brin, les conditions d'hybridation utilisées sont essentiellement similaires à celles décrites ci-dessus, et adaptables par l'homme du métier. On préfère dans ce cas procéder à l'hybridation en présence de formamide et on expose les complexes à une gamme de températures allant par exemple de 60 à 40 °C, préférentiellement de 56 °C à 44 °C, afin de favoriser la formation de complexes de type R-loop. De plus, il est souhaitable d'ajouter, après l'hybridation, un agent de stabilisation des triplex formés, une fois la formamide retirée du milieu, tels que le glyoxal par exemple (Kaback et.al. (1979) Nuc. Acid Res., 6, 2499-2517).

Ces hybridations croisées selon l'invention génèrent ainsi des compositions comprenant des hétéroduplex ou hétérotriplex ADNc/ARN, représentant les qualités de chacune des situations physiologiques testées. Comme indiqué ci-avant, dans chacune de ces compositions, des acides nucléiques correspondant essentiellement à des épissages alternatifs différentiels ou à d'autres altérations génétiques, spécifiques de chaque situation physiologique, peuvent être identifiés et/ou clonés.

L'invention concerne donc avantageusement un procédé

d'identification et/ou de clonage de régions d'acides représentatives de différences génétiques entre deux situations physiologiques, comprenant une étape d'hybridation entre les ARN provenant d'un échantillon biologique dans une première situation physiologique et les ADNc simple-brins  
5 provenant d'un échantillon biologique dans une deuxième situation physiologique, et l'identification et/ou le clonage, à partir des hybrides ainsi formés, des régions d'ARN non-appariées.

Cette première variante repose plus particulièrement sur la formation d'hétéroduplex entre les ARN et les ADNc simple-brin (voir Figures  
10 2-4). Cette variante est avantageusement mise en oeuvre en utilisant des ARN messagers ou des ADNc produits par transcription inverse des ARN messagers essentiellement, c'est-à-dire en présence d'une amorce oligo dT.

Dans un mode particulier de mise en oeuvre, le procédé d'identification et/ou de clonage d'acides nucléiques de l'invention  
15 comprend :

(a) l'hybridation d'ARN provenant de la situation test avec les ADNc simple-brin provenant de la situation de référence;

(b) l'hybridation d'ARN provenant de la situation de référence avec les ADNc simple-brin provenant de la situation test; et

20 (c) l'identification et/ou le clonage, à partir des hybrides formés en (a) et (b), de régions d'ARN non-appariées.

Dans une variante particulière de mise en oeuvre, le procédé de l'invention comprend les étapes suivantes :

25 (a) l'obtention d'ARN à partir d'un échantillon biologique dans une situation physiologique A (rA);

(b) l'obtention d'ARN à partir d'un même échantillon biologique dans une situation physiologique B (rB);

(c) la préparation d'ADNc à partir d'une partie des ARN rA obtenus en (a) (ADN cA) et à partir d'une partie des ARN rB obtenus en (b)  
30 (ADN cB) au moyen d'amorces polyT,

(d) l'hybridation en phase liquide d'une partie des ARN rA avec



une partie des ADN cB (pour générer des hétéroduplex rA/cB)

(e) l'hybridation en phase liquide d'une partie des ARN rB avec une partie des ADN cA (pour générer des hétéroduplex rB/cA),

(f) l'identification et/ou le clonage de régions d'ARN non appariées dans les hétéroduplex rA/cB et rB/cA obtenus en (d) et en (e).

Dans une autre variante particulière de mise en oeuvre, le procédé de l'invention comprend l'hybridation d'ARN provenant de la situation test avec les ADNc double-brins provenant de la situation de référence, et l'identification et/ou le clonage des régions d'ADN double-brin maintenues. Cette deuxième variante repose plus particulièrement sur la formation d'hétérotriplex entre les ARN et les ADNc double-brin, dérivé des structures de type R-loop (voir Figure 5). Cette variante est également préférentiellement mise en oeuvre en utilisant des ARN messagers ou des ADNc produits par transcription inverse des ARN messagers essentiellement, c'est-à-dire en présence d'une amorce polyT. Dans cette variante également, un mode de réalisation particulier comprend deux hybridations parallèles, générant deux populations d'acides nucléiques selon l'invention. Dans cette variante, les régions recherchées, spécifiques des épissages alternatifs, ne sont pas les régions d'ARN non appariées, mais des ADN double-brins qui n'ont pu être déplacés par une séquence ARN homologue (voir Figure 5).

Dans une autre variante de l'invention, le procédé comprend, pour isoler les différences génétiques qualitatives (e.g., les différences d'épissage) qui existent entre deux échantillons, l'hybridation entre une population d'ADNc double-brin provenant d'un premier échantillon biologique et une population d'ADNc (double-brin ou, préférentiellement simple-brin) provenant d'un deuxième échantillon biologique (figure 6).

A la différence des variantes présentées précédemment, celle-ci n'utilise pas des hétéroduplexes ou des hétérotriplexes ADN/ARN mais des homoduplexes ADN/ADN. Cette variante est avantageuse puisqu'elle ne donne pas seulement accès aux exons et aux introns alternatifs mais

également, et à l'intérieur d'une même banque d'acides nucléiques, aux jonctions spécifiques créées par délétion d'un exon ou d'un intron. De plus les séquences dans une telle banque donnent accès aux séquences flanquantes des exons et introns alternatifs.

5 Pour les deux échantillons (i.e., conditions physiopathologiques) étudiés, les ARN cytosoliques polyA<sup>+</sup> sont extraits selon les techniques connues de l'homme de métier et décrites précédemment. Ceux-ci sont convertis en ADNc par l'action d'une transcriptase inverse dépourvue ou non d'activité RNase H intrinsèque, comme décrit précédemment. L'un de ces  
10 ADNc simple brin est ensuite converti en ADNc double-brin par amorçage à l'aide d'hexamères aléatoires et selon les techniques connues de l'homme de l'art. Pour l'une des situations étudiées nous disposons donc d'un ADNc simple-brin (appelé "driver") et pour l'autre situation d'un ADNc double-brin (appelé "tester"). Ces ADNc sont dénaturés par chauffage puis mélangés de  
15 telle façon que le driver est en excès par rapport au tester. Cet excès est choisi entre 1 et 50 fois, avantageusement 10 fois. Dans une expérience donnée, menée à partir de deux situations physiopathologiques, le choix de la situation qui donne le driver est arbitraire et ne doit pas influencer la nature des informations recueillies. En effet, comme dans le cas des approches  
20 précédemment présentées, la stratégie d'identification des différences qualitatives qui existent entre deux populations d'ARNm repose sur le clonage de ces différences présentes dans des messagers communs : la stratégie repose sur le clonage de séquences présentes au sein de duplex et non de simples brins correspondant à des séquences singulières ou en  
25 excès dans l'une des situations étudiées. Le mélange des populations d'ADNc est précipité puis repris dans une solution contenant du formamide (par exemple 80%). L'hybridation est menée de 16 heures à 48 heures, avantageusement 24 heures. Les produits de cette hybridation sont précipités puis soumis à l'action d'une endonucléase de restriction ayant un  
30 site de reconnaissance de l'ADN double-brin dicté par 4 bases. Une telle enzyme de restriction va donc cliver l'ADNc double brin formé lors de

l'hybridation en moyenne toutes les 256 bases. Cette enzyme est sélectionnée avantageusement pour générer des sites cohésifs. Des exemples de telles enzymes sont fournis par des enzymes de restriction telles Sau3AI, HpaII, TaqI et MseI. Sont donc accessibles à une stratégie de clonage utilisant les sites de restriction clivés les fragments double-brin digérés par ces enzymes. Ces fragments sont de deux types : des fragments parfaitement hybridés, dont les deux brins sont parfaitement complémentaires, et des fragments dont l'hybridation est partielle c'est à dire comprenant une boucle simple brin encadrée par des régions double-brin (Figure 6A). Ces derniers fragments, minoritaires, contiennent les informations d'intérêt. Afin de les séparer des fragments parfaitement hybridés, majoritaires puisque dérivés de la majorité de la longueur des ADNc, des techniques de séparation sur gel ou sur toute autre matrice appropriée sont utilisées. Ces techniques mettent à profit le retard de migration, électrophorétique ou lors de gel filtration notamment, des fragments d'ADN qui contiennent une boucle d'ADN simple-brin. Ainsi les populations de fragments minoritaires qui contiennent les informations désirées peuvent être séparées de façon préparative des populations de fragments majoritaires correspondant aux régions d'ADN identiques dans les deux populations. Cette variante, qui permet d'isoler au sein d'une même population les empreintes positives et négatives liées à des différences qualitatives, peut également être appliquée à des hétéroduplexes ARN/ADN. A cet égard, un exemple de retard de migration d'un hétéroduplexe ARN/ADN dans lequel une partie de l'ARN n'est pas appariée, par rapport à un hétéroduplexe homologue dans lequel toutes les séquences sont appariées est illustré sur le modèle grb2/grb33 décrit dans les exemples (voir notamment la figure 8, puits 2 et 3).

#### Identification et/ou Clonage

A partir des populations d'acides nucléiques générées par hybridation, les régions caractéristiques des différences qualitatives (e.g.,

des épissages alternatifs différentiels) peuvent être identifiées par toute technique connue de l'homme du métier.

Identification et/ou clonage à partir des hétéroduplexes

5    ARN/ADN

Ainsi, dans le cas des hétéroduplex ARN/ADN (première variante du procédé), ces régions se présentent essentiellement sous forme de régions d'ARN non-appariées (boucles d'ARN), comme représenté sur la Figure 3. Ces régions peuvent donc être identifiées et clonées par séparation  
10 des hétéroduplexes, et des acides nucléiques simple-brin (ADN, ARN) (excès d'acide nucléique n'ayant pas réagi), digestion sélective des ARN double-brins (domaines engagés dans les hétéroduplex), puis séparation des ARN simple-brin résultant et des ADN simple-brins.

A cet égard, selon une première approche illustrée sur la Figure  
15 3, les régions d'ARN non appariées sont identifiées par traitement des hétéroduplex au moyen d'une enzyme capable de digérer sélectivement les domaines des ARN engagés dans des hétéroduplex ARN/ADN. Des enzymes douées de cette propriété sont décrites dans l'art antérieur et sont disponibles dans le commerce. Ce sont les RNases H, telles que en  
20 particulier, celle de E. Coli produite sous forme recombinante et disponible dans le commerce (Promega Réf. M4281 ; Life Technologies Réf. 18021). Ce premier traitement génère donc un mélange comprenant les régions d'ARN non appariées simple-brin et les ADNc simple-brin. Les ARNs peuvent être séparés des ADNc par toute technique connue de l'homme du métier, et  
25 notamment sur la base du marquage des amorces utilisées pour la préparation des ADNc (voir ci-dessus). Ces ARN peuvent être utilisés comme source de matériel pour l'identification de cibles, de produits génétiques d'intérêt ou toute autre application. Ces ARN peuvent également être convertis en ADNc, puis clonés dans des vecteurs, comme décrit ci-  
30 après.

A cet égard, le clonage des ARNs peut être réalisés de

différentes façons. L'une consiste à insérer à chaque extrémité des ARN des oligonucléotides servant de matrice à une réaction de transcription inverse en présence des amorces correspondantes. Cet ajout d'amorces se fait selon les techniques bien connues de l'homme du métier grâce à une enzyme, telle que par exemple la ARN ligase qui provient du phage T4 et qui catalyse la formation de liaisons phosphodiester intermoléculaires entre le phosphate en 5' d'une molécule donneuse et l'hydroxyl en 3' d'une molécule acceptrice. Une telle ARN ligase est disponible commercialement (par exemple Life Technologies - GIBCO BRL Réf. 18003). Les ADNc ainsi obtenus peuvent ensuite être amplifiés par les techniques classiques (PCR par exemple) en utilisant les amorces appropriées, comme illustré sur la Figure 3. Cette technique est particulièrement adaptée au clonage des ARN de petite taille (inférieure à 1000 b).

Une autre approche pour le clonage et/ou l'identification des régions d'ARN spécifiques consiste par exemple à réaliser une transcription inverse, sur le produit de digestion par une enzyme spécifique des ARN engagés dans des double-brins, telle la Rnase H, en utilisant des amorces aléatoires, qui vont initier la transcription au hasard à l'intérieur des ARNs. Les ADNc obtenus sont ensuite amplifiés selon les techniques classiques de biologie moléculaire, par exemple par PCR en utilisant des amorces grâce à des oligonucléotides ajoutés aux extrémités des ADNc grâce à l'action de l'ARN ligase du phage T4 (disponible commercialement ; par exemple chez Life Technologies - GIBCO BRL ref. 18003). Cette seconde technique est illustrée sur la Figure 4 et dans les exemples. Cette technique est plus particulièrement adaptée aux ARNs de taille importante, et permet d'obtenir une partie de l'information de séquence, suffisante pour reconstituer par la suite la totalité de la séquence de départ.

Une autre approche pour le clonage et /ou l'identification des régions d'ARN spécifiques repose également sur la réalisation d'une transcription inverse en utilisant des amorces aléatoires (figure 4). Néanmoins, selon cette variante, les amorces utilisées sont au moins en

partie des amorces semi-aléatoires, c'est-à-dire des oligonucléotides comprenant:

- une région aléatoire (de dégénérescence),
- une zone minimale d'amorçage présentant un degré de
- 5 contrainte défini, et
- une zone stabilisatrice.

De préférence, il s'agit d'oligonucléotides comprenant, dans l'orientation 5' → 3' :

- une zone stabilisatrice comprenant 8 à 24 nucléotides
- 10 déterminés, de préférence de 10 à 18 nucléotides. Cette zone stabilisatrice peut elle-même correspondre à la séquence d'un oligonucléotide utilisé pour réamplifier les fragments issus des premières amplifications réalisées à l'aide des amorces semialéatoires de l'invention. En outre, la zone stabilisatrice peut comprendre la séquence d'un ou plusieurs sites, de préférence non-
- 15 palindromiques, correspondant à des enzymes de restriction. Ceci permet par exemple de faciliter le clonage des fragments ainsi amplifiés. Un exemple particulier de zone stabilisatrice est représenté par la séquence GAG AAG CGT TAT (résidus 1 à 12 de SEQ ID NO:1);

- une région aléatoire ayant de 3 à 8 nucléotides, plus
- 20 particulièrement de 5 à 7 nucléotides, et

- une zone minimale d'amorçage définie de sorte que l'oligonucléotide s'hybride en moyenne au moins toutes les 60 pb environ, de préférence toutes les 250 pb environ. Plus préférentiellement, la zone d'amorçage comporte de 2 à 4 nucléotides définis, préférentiellement 3 ou 4,
- 25 tels que par exemple AGGX, où X représente l'une des quatre bases A, C, G ou T. La présence d'une telle zone d'amorçage confère à l'oligonucléotide la capacité d'hybrider en moyenne toutes les 256 paires de bases environ.

De manière particulièrement préférée, il s'agit d'oligonucléotides de formule :

- 30 GAGAAGCGTTATNNNNNNNAGGX (SEQ ID NO: 1) où les bases fixées ont été ordonnées de façon à minimiser le bruit de fond dû à

des auto-appariements dans des expériences de PCR, où N indique que les quatre bases peuvent être présentes de façon aléatoire à la position indiquée, et où X représente l'une des bases A, C, G ou T. De tels oligonucléotides constituent également un objet de la présente invention.

5 A cet égard, de façon à augmenter les possibilités d'amorçage sur les ARN à cloner, des réactions en parallèle peuvent être effectuées avec des oligonucléotides tels que :

GAGAAGCGTTATNNNNNNNAGGT (oligonucléotides A)

GAGAAGCGTTATNNNNNNNAGGA (oligonucléotides B)

10 GAGAAGCGTTATNNNNNNNAGGC (oligonucléotides C)

GAGAAGCGTTATNNNNNNNAGGG (oligonucléotides D),

chaque population d'oligonucléotides (A, B, C, D) pouvant être utilisée individuellement ou en combinaison avec une autre.

Après l'étape de transcription inverse, les ADNc sont amplifiés  
15 par PCR en utilisant les oligonucléotides A ou B ou C ou D.

Comme indiqué ci-avant, selon la complexité et la spécificité de la population des oligonucléotides souhaitée le nombre de positions dégénérées peut varier de 3 à 8, de préférence de 5 à 7. En deçà de 3 les hybridations sont restreintes et au delà de 8 la population d'oligonucléotides  
20 est trop complexe pour assurer une bonne amplification de bandes spécifiques.

Par ailleurs, la longueur de l'extrémité 3' fixe (la zone d'amorçage contrainte) de ces oligonucléotides peut également être modifiée : si les amorces décrites plus haut, avec 4 bases fixées, permettent  
25 d'amplifier en moyenne des fragments de 256 paires de bases, des amorces avec 3 bases fixes permettent d'amplifier des fragments plus courts (64 paires de bases en moyenne). Dans un premier mode préféré de l'invention, on utilise des oligonucléotides dans lesquels la zone d'amorçage comprend 4 bases fixes. Dans un autre mode préféré de l'invention, on utilise des  
30 oligonucléotides ayant une zone d'amorçage de trois bases fixes. En effet, les exons ayant une taille moyenne de 137 bases, ceux-ci sont

avantageusement amplifiés avec de tels oligonucléotides. A cet égard, voir également les oligonucléotides de séquence SEQ ID NO: 2, 3 et 4, par exemple.

Enfin, généralement, l'étape d'identification et/ou de clonage  
5 des ARN met en oeuvre les différentes méthodes de PCR et de clonage, de manière à obtenir l'information la plus complète.

. Identification et/ou clonage à partir des hétérotriplexes.

Dans le cas des hétérotriplexes (autre variante du procédé), les  
10 régions de différences qualitatives (insertions, délétions, épissages différentiels) se présentent essentiellement sous forme de régions d'ADN double-brin, comme représenté sur la Figure 5. Ces régions peuvent donc être identifiées et clonées par traitement en présence d'enzymes appropriées telles qu'une enzyme permettant de digérer les ARN, puis une enzyme  
15 permettant de digérer les ADN simple-brin. Les acides nucléiques ainsi obtenus sont donc directement sous forme d'ADN double-brin et peuvent être clonés dans tout vecteur approprié, tel le vecteur pMos-Blue (Amersham, RPN 5110), par exemple. Cette méthodologie est à différencier des approches déjà décrites utilisant des ARNs ou oligonucleotides de  
20 séquences prédéterminées, modifiés pour exercer une activité nucléasique (Landgraf et al. (1994) Biochemistry, 33, 10607-10615).

. Identification et/ou clonage à partir des homoduplexes ADN/ADN (figure 6).

25 Les fragments isolés par leurs structures atypiques sont ensuite additionnés, à chacune de leurs extrémités, d'adaptateurs, ou linkers, ayant des sites de restriction clivés à l'une de leurs extrémités. Cette étape peut être réalisée selon les techniques connues de l'homme du métier, par exemple par ligation avec l'ADN ligase du phage T4. Les sites de restriction  
30 ainsi introduits sont choisis compatibles avec les sites des fragments d'ADNc. Les linkers introduits sont des séquences d'ADNc double-brin, de



séquences connues, permettant de dériver des amorces pour réaliser des amplifications enzymatiques (PCR). Puisque l'étape suivante consiste à amplifier les deux brins qui présentent entre eux les différences qualitatives à identifier, il est nécessaire d'utiliser des linkers dont les extrémités 5' sont phosphorylées. Ainsi après dénaturation thermique des ADNc double-brin additionnés de linkers, chacune des extrémités de ces ADNc est liée de façon covalente avec une séquence d'amorçage spécifique. Après PCR à l'aide des amorces spécifiques appropriées, deux catégories d'ADNc double-brin sont obtenues: des fragments qui contiennent des séquences spécifiques de différences qualitatives qui distinguent les deux situations physiopathologiques, et des fragments qui comprennent l'empreinte négative de ces événements d'épissages. Le clonage de ces fragments permet d'obtenir une banque d'épissages alternatifs dans laquelle, pour chaque événement d'épissage, des empreintes positives et négatives sont présentes. Dans cette banque, sont donc accessibles non seulement les exons et les introns alternatifs mais aussi les jonctions spécifiques créées par excision de ces séquences épissées. Dans une même banque, ces différentes informations génétiques peuvent provenir des deux situations physiopathologiques sans discrimination. Par ailleurs, de façon à vérifier le caractère différentiel des épissages identifiés et de façon à déterminer dans quelle situation ceux-ci sont spécifiquement recrutés, les clones de la banque peuvent être hybridés avec des sondes dérivées de chacune des populations totales des ARNm.

Deux usages principaux peuvent être envisagés pour les fragments d'ADNc issus des différences qualitatives identifiées :

- Leur clonage dans des vecteurs appropriés de façon à constituer des banques représentatives des différences qualitatives qui existent entre les deux situations physiopathologiques étudiées,
- Leur utilisation en tant que sondes afin de cribler une banque d'ADN permettant d'identifier les événements épissés de façon différentielle.

Les vecteurs utilisés dans l'invention peuvent être notamment des plasmides, cosmides, phages, YAC, HAC, etc. Ces acides nucléiques peuvent ainsi être conservés tels quels ou introduits dans des microorganismes adaptés au vecteur de clonage utilisé, afin d'être multipliés et/ou conservés sous forme de cultures.

Les méthodes telles que décrites ci-dessus sont généralement mises en oeuvre, pour chaque échantillon, sur une période de temps de moins de deux mois, en particulier moins de 6 semaines. Par ailleurs, ces différentes méthodes peuvent être automatisées afin de réduire la durée totale et de faciliter le traitement de nombreux échantillons.

A cet égard, un autre objet de l'invention concerne les acides nucléiques identifiés et/ou clonés par les méthodes de l'invention. Comme indiqué ci-dessus, ces acides nucléiques peuvent être des ARN ou des ADNc. Plus généralement, l'invention concerne une composition d'acides nucléiques, comprenant essentiellement des acides nucléiques correspondant aux épissages alternatifs distinguant deux situations physiologiques. Plus particulièrement, ces acides nucléiques correspondent aux épissages alternatifs identifiés dans un échantillon biologique test et non présents dans le même échantillon biologique dans une situation de référence. L'invention a également pour objet l'utilisation des acides nucléiques ainsi clonés comme produit thérapeutique ou diagnostic, ou comme outil de criblage de molécules actives, comme indiqué ci-après.

Les différentes méthodes exposées ci-dessus aboutissent donc toutes au clonage de séquences d'ADNc qui représentent l'information génétique épissée différenciellement entre deux situations physiopathologiques. L'ensemble des clones issus de l'une de ces méthodes permet donc la constitution d'une banque représentative des différences qualitatives qui existent entre deux situations étudiées.

#### Génération de banques qualitatives

A cet égard, l'invention concerne en outre un procédé de

préparation d'une banque d'acides nucléiques représentatifs d'un état physiologique donné d'un échantillon biologique. Ce procédé comprend avantageusement le clonage d'acides nucléiques représentatifs des marqueurs qualitatifs d'expression génétiques (par exemple des épissages alternatifs) dudit état physiologique et non présents dans un état de référence, dans des banques spécifiques de différences qualitatives qui existent entre les 2 états étudiés.

Ces banques sont constituées d'ADNc insérés dans des vecteurs plasmidiques ou phagiques. Ces banques peuvent être présentées sur des filtres de nitrocellulose ou tout autre support connu de l'homme de l'art, tels des chips ou biopuces.

L'une des caractéristiques et en même temps l'une des originalités du criblage différentiel qualitatif est que cette technique aboutit à non pas une mais avantageusement deux banques différentielles qui représentent l'ensemble des différences qualitatives qui existent entre deux situations données : Paire de banque (voir figure 1D).

Ainsi, l'invention concerne de préférence toute composition ou banque d'acides nucléiques, susceptible d'être obtenue par hybridation entre une population d'ARN provenant d'un premier échantillon biologique et une population d'ADNc provenant d'un deuxième échantillon biologique. Plus préférentiellement, les banques ou compositions de l'invention comprennent des acides nucléiques représentatifs des différences qualitatives d'expression entre deux échantillons biologiques, et sont produites par un procédé comprenant (i) une étape au moins d'hybridation entre une population d'ARN provenant d'un premier échantillon biologique et une population d'ADNc provenant d'un deuxième échantillon biologique, (ii) la sélection des acides nucléiques représentatifs des différences qualitatives d'expression et, éventuellement (iii) le clonage desdits acides nucléiques.

En outre, après constitution de telles banques, il est possible d'effectuer une étape de sélection des clones pour améliorer la spécificité des banques obtenues. En effet, il est possible que certains

mésappariements observés ne soient pas uniquement dus à différences qualitatives (e.g. à des épissages alternatifs différentiels), mais puissent résulter de défaut(s) de la transcription inverse par exemple. Bien que ces événements ne soient pas généralement significatifs, il est préférable de les éliminer ou de les réduire préalablement au clonage des acides nucléiques. Pour ce faire, les clones de la banque peuvent être hybridés avec les populations d'ADNc des deux situations physiologiques étudiées (voir étape (c) ci-dessus). Les clones hybridant de façon non différentielle avec les deux populations peuvent être considérés comme non-spécifiques et éventuellement éliminés ou traités en deuxième priorité (en effet, l'apparition d'une nouvelle isoforme dans l'échantillon test ne signifie pas toujours que l'isoforme initiale présente dans l'échantillon de référence a disparu de cet échantillon test). Les clones n'hybridant qu'avec une seule des deux populations ou hybridant de façon préférentielle avec l'une des populations sont considérés comme spécifiques et peuvent être sélectionnés en première priorité pour constituer des banques enrichies ou affinées.

Un affinage peut également être réalisé par hybridation et validation de clones avec des sondes provenant d'un nombre statistiquement relevant d'échantillons pathologiques.

La présente demande a donc également pour objet toute banque d'acides nucléiques comprenant des acides nucléiques spécifiques d'épissages alternatifs caractéristiques d'une situation physiologique. Ces banques sont avantageusement constituées d'ADNc, généralement double brin, correspondant aux régions d'ARN spécifiques d'un épissage alternatif. Ces banques peuvent être constituées des acides nucléiques, généralement dans un vecteur de clonage ou de cultures cellulaires contenant lesdits acides nucléiques.

Le choix des ARN de départ détermine en partie les caractéristiques des banques obtenues :

- les ARN des deux situations A et B sont des ARNm ou des ARN totaux matures isolés selon les techniques connues de l'homme de l'art.

Les banques sont alors des banques de criblage différentiel qualitatif dites restreintes, car restreintes aux différences qualitatives qui caractérisent les ARN matures des deux situations physiopathologiques.

- Les ARN de l'une des situations sont des ARNm ou totaux matures alors que les ARN de l'autre situation sont des ARN pré-messagers, non maturés par épissage, isolés selon les techniques connues de l'homme de l'art, à partir de noyaux cellulaires. Dans ce cas les banques obtenues sont des banques de criblage différentiel dites complexes, puisque non restreintes aux différences entre ARN matures mais comprenant tout le répertoire des épissages transcrits dans une situation et éliminés dans l'autre, dont tous les introns.

enfin, les ARN peuvent provenir d'une seule situation physiopathologique et dans ce cas le criblage différentiel implique les ARN matures et les pré-messagers d'un même échantillon. Dans ce cas, les banques obtenues sont des banques de criblage différentiel qualitatif autologues. L'intérêt de telles banques est qu'elles rassemblent exclusivement le répertoire des introns transcrits dans une situation donnée. Leur hybridation avec une sonde provenant d'ARN matures d'une autre situation détermine rapidement si à cette situation est caractérisée par une rétention d'introns tout en permettant aisément leur identification.

Généralement, les banques sont générées par étalement, sur milieu solide (notamment sur milieu gélosé), d'une culture cellulaire transformée par les acides nucléiques clonés. La transformation est réalisée par toute technique connue de l'homme du métier (transfection, phosphate de calcium, électroporation, infection par des bactériophages, etc). La culture cellulaire est généralement une culture de bactéries, telles que par exemple les bactéries E. coli. Il peut également s'agir de cultures de cellules eucaryotes, notamment de cellules eucaryotes inférieures (levures par exemple). Cet étalement peut être réalisé sur boîte ou sur tout autre support adapté, en conditions stériles. En outre, ces cultures étalées en milieu gélosé peuvent être stockées sous forme congelée par exemple (dans du glycérol

ou autre agent adapté). Ces banques peuvent naturellement être utilisées pour la production de "répliques", c'est-à-dire de copies selon les techniques habituelles détaillées ci-après. En outre, ces banques servent généralement à préparer une banque amplifiée, c'est-à-dire une banque comprenant

5 chaque clone sous forme amplifiée. Une banque amplifiée est préparée comme suit : à partir de la culture étalée, tous les clones cellulaires sont récupérés et sont conditionnés pour être conservés sous forme congelée ou au froid, dans tout milieu adapté. Cette banque amplifiée est

10 avantageusement réalisée à partir de cultures de bactéries E.coli, et conservée à 4°C, en conditions stériles. Cette banque amplifiée permet la préparation et la reproduction illimitée de toute banque ultérieure contenant ces clones, sur différents supports, pour différentes applications. Une telle banque permet en outre l'isolement et la caractérisation de tout clone d'intérêt. Chacun des clones constituant les banques de l'invention est en

15 effet un élément caractéristique d'une situation physiologique, et constitue donc une cible particulièrement intéressante pour différentes études telles que la recherche de marqueurs, la préparation d'anticorps, le diagnostic, le traitement pour transfert de gènes, etc. Ces différentes applications sont discutées plus en détail plus loin. La banque est généralement préparée

20 comme décrit ci-dessus par étalement des cultures dans un milieu gélosé, sur un support adapté (boîte de pétri par exemple). L'intérêt d'utiliser un milieu gélosé est que chaque colonie peut être séparée et individualisée. A partir de cette culture, des répliques à l'identique peuvent être préparées en quantités importantes par simple "réplique" sur tout support approprié selon

25 les techniques de l'homme de l'art. Ainsi, la réplique peut être réalisée au moyen de filtres, membranes (nylon, nitrocellulose, etc) permettant l'accrochage des cultures. Les filtres peuvent ensuite être stockés en l'état, à 4°C par exemple, sous forme desséchée, dans tout type de conditionnement qui n'altère pas les acides nucléiques. Les filtres peuvent également être

30 traités de manière à éliminer les cellules, protéines, etc, et à ne conserver que des composants tels que les acides nucléiques. Ces traitement peuvent

comprendre notamment des protéases, des détergents, etc. Les filtres traités peuvent également être conservés dans tout dispositif ou toute condition adaptés aux acides nucléiques.

Les banques d'acides nucléiques peuvent également être  
5 préparées directement à partir des acides nucléiques, par dépôt sur des biopuces ou tout autre dispositif approprié.

L'invention concerne également toute banque comprenant des oligonucléotides spécifiques d'épissages alternatifs distinguant deux situations physiologiques. Il s'agit avantageusement d'oligonucléotides  
10 simple-brin, comprenant de 5 à 100-mères, de préférence moins de 50-mères, par exemple autour de 25-mères environ.

Ces oligonucléotides sont spécifiques d'épissages alternatifs représentatifs d'une situation ou d'un type de situation physiologique. Ainsi, de tels oligonucléotides peuvent être par exemple des oligonucléotides  
15 représentatifs d'événements d'épissages alternatifs caractéristiques de situations d'apoptose. Il a en effet été décrit dans la littérature que certains épissages alternatifs étaient observés dans le cadre de situations apoptotiques. Il s'agit par exemple d'épissages dans les gènes Bclx, Bax, Fas ou Grb2 notamment. A partir des données publiées et des séquences  
20 accessibles dans la littérature et/ou sur bases de données, il est possible de créer des oligonucléotides spécifiques des formes épissées et non épissées. Ces oligonucléotides peuvent par exemple être créés selon la stratégie suivante :

(a) identification d'une protéine ou d'un événement d'épissage  
25 caractéristique d'une situation d'apoptose et de la séquence du domaine épissé. Cette identification peut être basée sur des données publiées ou par compilation de séquences accessibles sur bases de données;

(b) synthèse artificielle d'un ou plusieurs oligonucléotides correspondant à une ou plusieurs régions de ce domaine, qui permettent  
30 donc par hybridation de mettre en évidence la forme non épissée dans les ARN d'un échantillon test;

(c) synthèse artificielle d'un ou plusieurs oligonucléotides correspondant à la région de jonction entre les deux domaines séparés par le domaine épissé. Ces oligonucléotides permettent donc par hybridation de mettre en évidence la forme épissée dans les ARN d'un échantillon test;

5 (d) reproduction des étapes (a) à (c) ci-dessus avec d'autres protéines ou événements d'épissages caractéristiques d'une situation d'apoptose;

(e) transfert sur un premier support approprié du ou des oligonucléotides spécifiques des formes apoptotiques des messagers  
10 identifiés ci-avant et, sur un autre support approprié, du ou des oligonucléotides spécifiques des formes non-apoptotiques.

Les deux supports ainsi obtenus peuvent être utilisés pour tester l'état physiologique de cellules ou échantillons tests, et notamment leur état apoptotique, par hybridation d'une préparation d'acides nucléiques de  
15 ces cellules ou échantillons.

D'autres banques similaires peuvent être générées avec des oligonucléotides spécifiques d'états physiopathologiques différents (neurodégénérescence, toxicité, prolifération, etc.) et ainsi permettre un élargissement des domaines d'applications.

20 Des banques d'introns ou d'exons alternatifs peuvent aussi être des banques de données informatiques constituées par analyse systématique des banques de données qui regroupent les informations relatives au génome de tel ou tel organisme, tissu ou culture cellulaire. Dans ce cas, les données obtenues par constitution de telles banques virtuelles  
25 peuvent être utilisées pour générer des amorces oligonucléotidiques qui seront utilisées pour tester en parallèle deux situations physiopathologiques.

Les données des banques informatiques peuvent également être utilisées pour dériver des sondes nucléotidiques générales, représentatives d'une classe de protéines ou encore spécifiques d'une  
30 séquence définie. Ces sondes peuvent ensuite être appliquées sur les banques de clones issues des différentes techniques de clonage des introns



et exons alternatifs afin d'obtenir une image de la complexité de ces banques moléculaires et de déterminer rapidement si telle ou telle classe de protéines ou telle ou telle séquence déterminée est épissée différemment entre deux états physiopathologiques distincts.

5                   Une autre banque ou compositions d'acides nucléiques selon l'invention est une banque antisens, réalisée à partir des séquences identifiées selon les méthodes de l'invention (DATAS). Pour la réalisation de ce type de banques, ces séquences sont clonées de façon à être exprimées en fragments d'ARN correspondant à une orientation antisens par rapport  
10 aux ARN messagers sur lesquels ont été réalisés DATAS. On aboutit ainsi à une banque dite antisens. Cette approche utilise de préférence la variante de clonage qui permet une orientation des fragments clonés. L'intérêt d'une telle banque antisens est de permettre la transfection de lignées cellulaires et de suivre l'altération de tout phénotype qu'il soit d'ordre morphologique, enzymatique ou suivi par l'utilisation de gènes rapporteurs ou de résistance à  
15 un agent de sélection. L'analyse des variations phénotypiques consécutives à l'introduction d'un vecteur d'expression antisens se fait généralement après sélection de clones dits stables, c'est à dire permettant une réplication coordonnée du vecteur d'expression et du génome de l'hôte. Cette  
20 coordination est permise par l'intégration du vecteur d'expression dans le génome cellulaire ou, lorsque le vecteur d'expression est épisomal, par pression de sélection. Cette pression de sélection se fait par traitement de la culture cellulaire transfectée avec un agent toxique qui ne peut être détoxifié que lorsque le produit d'un gène porté par le vecteur d'expression est  
25 exprimé dans la cellule. Il s'ensuit une synchronisation entre la réplication de l'hôte et celle du transgène. Avantageusement, on utilise des vecteurs épisomaux dérivés du virus Epstein-Barr qui permettent l'expression dans une même cellule de 50 à 100 copies du vecteur (Deiss et al, 1996, EMBO J., 15, 3861-3870 ; Kissil et al, 1995, J. Biol. Chem, 270, 27932-27936).

30                   L'intérêt de ces banques antisens alliées aux séquences DATAS qu'elles contiennent est d'identifier non seulement quel gène a eu

son expression inhibée pour amener au phénotype sélectionné mais aussi d'identifier quelle isoforme d'épissage de ce gène a été affectée. Lorsque le fragment antisens cible un exon donné, il peut en être déduit que le domaine protéique et donc la fonction impliquant ce domaine s'oppose au phénotype observé. En cela le couplage de DATAS avec une approche antisens représente un raccourci vers la génomique fonctionnelle.

#### Puces à ADN

L'invention concerne également tout support (membrane, filtre, biopuce, chip, etc) comprenant une banque ou une composition d'acides nucléiques telle que définie ci-dessus. Il peut s'agir plus particulièrement d'une banque cellulaire ou d'une banque d'acides nucléiques. L'invention concerne également tout kit ou support comprenant plusieurs banques selon l'invention. En particulier, il peut être avantageux d'utiliser en parallèle une banque représentative des qualités d'un état physiologique test par rapport à un état physiologique de référence et, à titre de contrôle, une banque représentative des qualités de l'état physiologique de référence par rapport à l'état physiologique test ("paire de banques"). Un kit avantageux selon l'invention comprend donc deux banques qualitatives différentielles de deux situations physiologiques (une "paire de banques"). Selon un mode de réalisation particulier, les kits de l'invention comprennent plusieurs paires de banques telles que définies ci-dessus, correspondant à différents états physiologiques ou à différents échantillons biologiques par exemple. Les kits peuvent comprendre par exemple ces différentes paires de banques déposées en série sur un même support.

#### Génération de sondes

Une autre utilisation des compositions d'ADNc selon l'invention, représentatifs des différences qualitatives qui existent entre deux états physiopathologiques, consiste à en dériver des sondes. De telles sondes peuvent en effet être utilisées pour cribler les événements épissés de façon

différentielle entre deux situations physiopathologiques.

Ces sondes (voir figure 1D) peuvent être préparées par marquage des populations ou banques d'acides nucléiques selon les techniques classiques, connues de l'homme du métier. Ainsi, il peut s'agir de marquage enzymatique, radioactif, fluorescent, immunologique, etc. Préférentiellement, il s'agit d'un marquage radioactif ou fluorescent. Ce type de marquage peut être réalisé par exemple en introduisant sur la population d'acides nucléiques (soit après synthèse soit au cours de leur synthèse) des nucléotides marqués, permettant leur révélation par les méthodes conventionnelles.

Une application est donc de cribler une banque génomique classique. Une telle banque peut comprendre selon le vecteur, dérivé d'un phage ou d'un cosmide, des fragments d'ADN de 10kb à 40kb. Le nombre de clones hybridant avec les sondes générées par DATAS et représentatifs des différences d'épissage qui existent entre deux situations reflète donc à peu près le nombre de gènes affectés par des variations d'épissage, selon qu'ils sont exprimés dans l'une ou l'autre des situations étudiées.

De préférence, les sondes de l'invention sont utilisées pour cribler une banque d'ADN génomique (généralement humaine) adaptée à l'identification d'événements d'épissage. De préférence, une telle banque génomique est composée de fragments d'ADN de taille restreinte (généralement clonés dans des vecteurs), de façon à statistiquement ne recouvrir qu'un seul élément épissable différenciellement, c'est à dire un seul exon ou un seul intron. La banque d'ADN génomique est donc préparée par digestion d'ADN génomique avec une enzyme ayant un site de reconnaissance restreint par 4 bases, assurant ainsi la possibilité d'obtenir par digestion ménagée des fragments d'ADN de taille moyenne de 1kb. De tels fragments nécessitent l'obtention de  $10^7$  clones pour constituer une banque d'ADN représentative d'un génome d'organisme eucaryote supérieur. Une telle banque constitue également un objet de la présente demande. Cette banque est ensuite hybridée avec les sondes dérivées du

criblage différentiel qualitatif. Précisément, de chaque expérimentation considérée et qui compare deux situations physiopathologiques A et B, deux sondes (paire de sondes) sont obtenues. Une sonde enrichie en événements d'épissages caractéristiques de la situation A et une sonde enrichie en marqueurs d'épissage de B. Les clones de la banque génomique qui hybrident préférentiellement avec l'une ou l'autre sonde portent des séquences préférentiellement épissées dans les situations physiopathologiques correspondantes.

Les méthodes de l'invention permettent ainsi l'identification systématique de différences qualitatives d'expression génique. Ces méthodes présentent de nombreuses applications, dans l'identification et/ou le clonage de molécules d'intérêt, en toxicologie, en pharmacologie ou encore en pharmacogénomique par exemple.

#### Applications

L'invention concerne donc également l'utilisation des méthodes, acides nucléiques ou banques décrite ci-dessus pour l'identification de molécules d'intérêt thérapeutique ou diagnostique. L'invention concerne plus particulièrement l'utilisation des méthodes, acides nucléiques ou banques décrite ci-dessus pour l'identification de protéines ou domaines protéiques affectés dans une pathologie.

L'un des atouts de ces techniques est en effet d'identifier à l'intérieur d'un messager, et par conséquent de la protéine correspondante, les domaines fonctionnels qui sont affectés dans une pathologie donnée. Cela permet d'assigner à un domaine donné une importance dans le développement ou le maintien d'un état pathologique. L'avantage immédiat de restreindre à un domaine précis d'une protéine l'impact d'une dérégulation pathologique est de proposer celui-ci comme une cible relevante pour un criblage de petites molécules à visée thérapeutique. Ces informations constituent également des clefs qui permettent de concevoir des polypeptides à activité thérapeutique délivrables par thérapie génique; Ces

polypeptides peuvent notamment être des anticorps simples chaînes dérivés d'anticorps neutralisants dirigés contre les domaines identifiés par les techniques décrites ci avant.

Plus spécifiquement, les méthodes selon l'invention permettent  
5 d'obtenir des molécules qui peuvent:

- être des séquences codantes qui dérivent d'exons alternatifs.
- correspondre à des séquences non codantes portées par des introns épissés différenciellement d'un état physiopathologique à un autre.

De ces deux points, différents enseignements peuvent être  
10 tirés.

Les épissages alternatifs d'exons qui différencient deux états physiopathologiques traduisent un niveau de régulation de l'expression génétique qui permet de moduler ( plus précisément d'abolir ou d'instaurer) une ou plusieurs fonctions d'une protéine donnée. Ainsi la plupart des  
15 domaines structuraux et fonctionnels (SH2, SH3, PTB, PDZ, et les domaines catalytiques de différentes enzymes...) étant codés par plusieurs exons contigus, deux configurations peuvent se présenter :

i) Les domaines sont tronqués dans la situation pathologique (Zhu, Q. et al, 1994, J. Exp. Med., vol 180, n°2, pp461-470);  
20 cela indique que les chemins de signalisation impliquant ces domaines doivent être restaurés dans un but thérapeutique.

ii) Les domaines sont maintenus au cours d'une pathologie alors qu'ils sont absents dans une situation saine; ces domaines peuvent être considérés comme des cibles de criblage de petites molécules  
25 chimiques destinées à antagoniser les signaux transduits par l'intermédiaire de ces domaines.

Les séquences épissées différenciellement peuvent correspondre à des régions non-codantes situées en 5' ou en 3' de la séquence codante ou à des introns intervenant entre deux exons codants.  
30 Dans les régions non codantes, ces épissages différentiels peuvent traduire une modification de la stabilité ou de la traductibilité du messager (Bloom, T.

J., and Beavo, J. A., 1995, Proc. Natl. Acad. Sci. USA, vol93, n° 24, pp 14188-14192; Ambartsumian, N. et al., 1995, Gene, vol 159, n° 1, pp 125-130). Ces phénomènes doivent alors être recherchés sur la base de ces informations et peuvent mettre en évidence que l'accumulation ou la  
5 disparition de la protéine correspondante la désigne ainsi comme cible d'intérêt. Lorsque la rétention d'un intron se produit dans une séquence codante, il s'ensuit le plus souvent une troncature de la protéine naturelle par introduction de codon stop dans la phase de lecture (Varesco, L., et al, 1994, Hum. Genet., vol 93, n°3, pp281-286; Canton, H., et al., 1996, Mol.  
10 Pharmacol., vol 50, n° 4, pp799-807, Ion, A., et al, 1996, Am. J. Hum. Genet., vol 58, n°6, pp1185-1191). Avant de rencontrer ce codon stop, il se produit généralement une lecture de quelques codons supplémentaires, ce qui aboutit à adjoindre à la partie déjà traduite une séquence spécifique, témoin protéique de l'épissage alternatif. Ces acides aminés  
15 supplémentaires peuvent être utilisés pour générer des anticorps spécifiques de la forme alternative caractéristique de la situation pathologique. Ces anticorps peuvent être ensuite utilisés comme outils de diagnostic. La protéine tronquée voit ses propriétés modifiées, voire altérées. Ainsi des enzymes peuvent être amputées de leur domaine catalytique ou de leur  
20 domaine régulateur, devenant inactives ou constitutivement activées. Des adaptateurs peuvent perdre leur capacité à connecter différents partenaires d'une cascade de signalisation(Watanabe, K. et al, 1995, J. Biol. Chem., vol 270, n°23, pp13733-13739). Les produits d'épissage des récepteurs peuvent aboutir à des récepteurs qui ont perdu leur capacité à lier leur ligand  
25 (Nakajima, T. et al, 1996, Life Sci., vol58, n°9, pp761-768) et peuvent également générer des formes de récepteurs solubles par relargage de leur domaine extracellulaire (Cheng J., 1994, Science, vol263, n° 5154, pp1759-1762). Dans ce cas, des tests diagnostiques peuvent être envisagés, basés sur la circulation dans les différents fluides physiologiques de forme soluble  
30 de récepteur à un ligand donné.

L'invention concerne plus particulièrement l'utilisation des

méthodes, acides nucléiques ou banques décrits ci-dessus pour l'identification de domaines antigéniques spécifiques de protéines impliquées dans une pathologie. L'invention concerne également l'utilisation des acides nucléiques, protéines ou peptides tels que décrits ci-avant pour le diagnostic  
5 de pathologies.

L'invention concerne également une méthode d'identification et/ou de production de protéines ou domaines protéiques impliqués dans une pathologie comprenant :

(a) l'hybridation des ARN messagers d'un échantillon  
10 pathologique avec les ADNc d'un échantillon sain, ou l'inverse, ou les deux en parallèle,

(b) l'identification, dans les hybrides formés, des régions correspondant aux différences qualitatives (non appariées (ARN) ou appariées (ADN double brin)), spécifiques de l'état pathologique par rapport  
15 à l'état sain,

(c) l'identification et/ou la production de la protéine ou domaine protéique correspondant à une ou plusieurs régions identifiées en (b).

Les régions identifiées correspondent généralement à des épissages différentiels, mais il peut également s'agir d'autres altérations  
20 génétiques telles que insertion(s) ou délétion(s), par exemple.

La ou les protéines ou domaines protéiques peuvent être isolés, séquencés, et utilisés dans des applications thérapeutiques ou diagnostiques, notamment pour la préparation d'anticorps.

A titre d'exemple plus spécifique, le criblage différentiel qualitatif  
25 de l'invention permet avantageusement de mettre en évidence des gènes suppresseurs de tumeurs. En effet, de nombreux exemples indiquent que l'un des modes d'inactivation des gènes suppresseurs au cours de la progression tumorale est une inactivation par modulation de l'épissage de formes alternatives.

30 Ainsi, dans les carcinomes pulmonaires à petites cellules, le gène de la protéine p130 qui appartient à la famille RB (protéine du

rétinoblastôme) est muté à un site consensus d'épissage. La conséquence de cette mutation est l'élimination de l'exon 2 et une non synthèse de protéine due à la présence d'un codon stop précoce. Cette observation a été la première à souligner l'importance des membres de la famille RB dans la tumorigénèse. De même, dans certains cancers du poumon non à petites cellules, le gène de la protéine p16INK4A, protéine qui est un inhibiteur des kinases cyclines-dépendantes cdk4 et cdk6 est muté dans un site donneur d'épissage. Le résultat de cette mutation est la production d'une protéine tronquée à demi-vie courte ce qui a pour conséquence l'accumulation des formes phosphorylées, inactives, de RB. Par ailleurs, WT1, le gène suppresseur de tumeur de Wilms, est transcrit en plusieurs ARN messagers générés par épissages alternatifs. Dans les cancers du sein, les proportions relatives des différents variants sont modifiées par rapport au tissu sain, fournissant des outils diagnostics et des pistes pour comprendre l'importance des différents domaines fonctionnels de WT1 dans la progression tumorale. Ce même phénomène de modification des rapports entre différentes formes d'ARN messagers et d'isoformes protéiques lors de la transformation cellulaire est retrouvé pour la neurofibrine NF1. En outre, cette notion de modulation des phénomènes d'épissage qui signe la progression tumorale est soutenue également par l'exemple de HDM2 dont 5 épissages alternatifs sont détectés dans les carcinomes ovariens et pancréatiques et dont les expressions augmentent selon le stade d'avancement tumoral. D'autre part, dans les cancers de la tête et du cou, l'un des mécanismes d'inactivation de p53 implique une mutation dans un site consensus d'épissage.

Ces quelques exemples listés illustrent tout l'intérêt des techniques de l'invention basées sur la recherche systématique des variations d'épissage qui distinguent une tumeur donnée du tissu sain voisin. Les résultats qui en découlent permettent en effet non seulement la caractérisation de gènes suppresseurs de tumeurs déjà connus mais également, compte tenu de l'aspect original et systématique des techniques de criblage différentiel qualitatif, l'identification de nouvelles variations



d'épissages spécifiques de tumeurs affectant vraisemblablement de nouveaux gènes suppresseurs de tumeurs.

L'invention a donc également pour objet un procédé d'identification et/ou de clonage de gènes suppresseurs de tumeurs ou d'altérations génétiques (e.g., d'épissages) au sein de gènes suppresseurs de tumeurs, tel que défini ci-avant. Ce procédé peut avantageusement comprendre les étapes suivantes :

- (a) l'hybridation des ARN messagers d'un échantillon de tumeur avec les ADNc d'un échantillon sain, ou l'inverse, ou les deux en parallèle,
- 10 (b) l'identification, dans les hybrides formés, des régions spécifiques de l'échantillon tumoral par rapport à l'état sain,
- (c) l'identification et/ou le clonage de la protéine ou domaine protéique correspondant à une ou plusieurs régions identifiées en (b).

Les propriétés suppresseur de tumeur des protéines ou domaines identifiés peuvent ensuite être testées dans différents modèles connus. Ces protéines, ou leur forme native (possédant l'épissage observé dans le tissu sain), peuvent ensuite être utilisées dans des applications thérapeutiques ou diagnostiques, notamment en thérapie génique antitumorale.

20 La présente demande a donc également pour objet non seulement les différents aspects de mise en oeuvre de la technologie mais aussi l'exploitation des informations qui en découlent à des fins de recherche, de développement de criblage de petites molécules chimiques, de développement d'outils de thérapie génique ou de diagnostic.

25 A ce titre, l'invention concerne également l'utilisation des méthodes, acides nucléiques ou banques décrits ci-dessus en génotoxicologie, c'est-à-dire pour anticiper (prédire) le potentiel toxique de composés tests.

30 Les programmes génétiques engagés lors du traitement de cellules ou de tissus par des agents toxiques sont en grande partie corrélés aux phénomènes d'apoptose ou mort cellulaire programmée. L'importance

des phénomènes d'épissage alternatif dans la régulation de ces voies apoptotiques est bien illustrée par la littérature. Cependant, aucune technologie génomique décrite jusqu'à présent ne permettait de rechercher systématiquement et d'isoler de manière exhaustive les variations de séquences dues à des épissages alternatifs et distinguant deux situations physiopathologiques données. Les techniques de criblage différentiel qualitatif développées dans la présente invention permettent de regrouper l'ensemble des différences d'épissage qui existent entre deux situations dans des banques d'ADNc. La comparaison des séquences d'ARN (par exemple les ARN messagers) d'un tissu (ou d'une culture cellulaire) traité ou non avec un composé toxique de référence permet d'établir des banques d'ADNc qui regroupent les différences qualitatives de l'expression génique qui caractérisent l'action toxique étudiée. Ces banques d'ADNc peuvent ensuite être hybridées avec des sondes dérivées d'ARN extrait des mêmes tissus ou cellules traités avec un produit chimique dont on veut évaluer le potentiel toxique. La plus ou moins grande capacité de ces sondes à s'hybrider avec les informations génétiques spécifiques d'une situation toxique de référence permet de lui assigner un potentiel toxique. Par ailleurs, outre l'application de DATAS à la génération et l'utilisation de banques de différences qualitatives induites par des agents toxiques, une partie de l'invention consiste également à démontrer que des dérégulations dans l'épissage de certains ARNs messagers peuvent être induites par certains agents toxiques, à des doses inférieures aux IC50 mesurées dans des tests de cytotoxicité et d'apoptose connus de l'homme de l'art. De telles dérégulations (ou dysrégulations) peuvent être utilisées comme marqueurs pour le suivi de la toxicité et/ou de l'efficacité de molécules (chimiques ou génétiques).

L'invention concerne donc également toute méthode de détection ou de suivi du potentiel toxique et/ou thérapeutique d'un composé basée sur la détection de formes et/ou de profils d'épissages induits par ce composé sur un échantillon biologique. Elle concerne en outre l'utilisation de toute modification de formes et/ou de profils d'épissages comme marqueur

pour le suivi de la toxicité et/ou de l'efficacité de molécules.

L'évaluation ou le suivi du potentiel toxique peut s'effectuer plus particulièrement selon deux approches :

Selon la première approche, le criblage différentiel qualitatif  
5 peut être réalisé entre un tissu ou une culture cellulaire de référence, d'une part non traité et d'autre part traité par le produit dont on veut évaluer la toxicité. L'analyse des clones qui représentent les différences qualitatives spécifiquement induites par le produit permet ensuite éventuellement de détecter dans ces clones des événements caractéristiques d'ADNc impliqués  
10 dans les phénomènes liés à la toxicité comme l'apoptose.

L'apparition de ces marqueurs est suivie selon la dose et la durée du traitement par le produit et permet une approche de son profil toxicologique.

La présente demande a donc également pour objet un procédé  
15 d'identification, par criblage différentiel qualitatif selon les techniques présentées ci-dessus, de marqueurs de toxicité induits dans un système biologique modèle par un composé chimique dont on veut tester le potentiel toxique. A cet égard, l'invention concerne notamment un procédé d'identification et/ou de clonage d'acides nucléiques spécifiques d'un état  
20 toxique d'un échantillon biologique donné comprenant la préparation de banques différentielles qualitatives entre les ADNc et les ARN de l'échantillon après ou sans traitement par un composé toxique test, et la recherche de marqueurs de toxicité spécifiques des qualités de l'échantillon après traitement.

25 Selon la deuxième approche, des abaques pour différentes classes de produits toxiques rassemblent leur profil de toxicité selon les doses employées et selon les durées des traitements pour un tissu ou un modèle cellulaire de référence. A chaque point de ces abaques, des banques d'ADNc caractéristiques des différences génétiques qualitatives  
30 peuvent être établies. Ces banques sont des banques différentielles qualitatives, i.e., elles sont obtenues par extraction des informations

génétiq ues du point choisi dans les abaques et du point correspondant au modèle tissulaire ou cellulaire contrôle. Comme cela est illustré dans les exemples, le criblage différentiel qualitatif repose sur l'hybridation d'ARNm extraits d'une situation avec les ADNc issues d'une autre. Comme indiqué  
5 plus haut, le criblage différentiel qualitatif peut être également mené à partir d'ARN totaux ou d'ARN nucléaires contenant les prémessagers.

A cet égard, l'invention concerne un procédé de détermination ou d'évaluation de la toxicité d'un composé test sur un échantillon biologique donné comprenant l'hybridation :

- 10 - de banques différentielles entre les ADNc et les ARN dudit échantillon biologique à l'état sain et à un ou différents stades de toxicité résultant d'un traitement dudit échantillon avec un composé toxique de référence, avec,
  - une préparation d'acides nucléiques de l'échantillon biologique  
15 traité par ledit composé test, et
  - l'évaluation du potentiel toxique du composé test par analyse du degré d'hybridation avec les différentes banques.

Selon ce procédé, pour chaque situation (doses de composé et/ou temps d'incubation), deux hybridations réciproques sont avantageusement  
20 réalisées, entre :

- les ARN de la situation A (test) et les ADNc de la situation B (référence) (rA/cB)
- les ARN de la situation B (référence) et les ADNc de la situation A (test) (rB/cA).

25 A chaque situation toxique de référence, à chaque point des abaques, correspondent donc deux banques de criblage différentiel qualitatif. L'une de ces banques regroupe les variations qualitatives, c'est à dire notamment les épissages alternatifs, spécifiques de la situation normale de référence alors que l'autre banque rassemble les épissages spécifiques des  
30 événements toxiques. Ces banques sont répliquées sur des supports solides tels des filtres de nylon ou de nitrocellulose ou avantageusement sur des

chips. Ces banques formées initialement par des fragments d'ADNc de longueur variable (selon les événements d'épissages concernés) peuvent être optimisées par l'emploi d'oligonucléotides dérivés des séquences isolées initialement.

5                   Lorsqu'un composé chimique est proposé pour un développement pharmaceutique, il peut être appliqué sur les mêmes modèles tissulaires ou cellulaires que ceux répertoriés dans les abaques de toxicité. Des sondes moléculaires peuvent ensuite être réalisées à partir d'ARNm extraits des échantillons biologiques traités par le composé chimique d'intérêt. Ces sondes sont ensuite hybridées à des filtres portant l'ADNc des banques rA/cB et rB/cA. Par exemple, la banque rA/cB peut contenir les séquences spécifiquement présentes dans la situation normale et la banque rB/cA les éléments d'épissage alternatifs spécifiques de la situation de toxicité. L'innocuité ou la toxicité du composé chimique est alors  
10 aisément évaluée en fonction des profils d'hybridation d'une sonde dérivée d'ARNm extraits du modèle tissulaire ou cellulaire de référence traité par le composé testé:

- une hybridation efficace avec la banque rA/cB et aucun signal sur la banque rB/cA révèle une absence de toxicité du composé sur le  
20 modèle étudié

- l'hybridation de la sonde avec des clones de la banque rB/cA indique une toxicité induite par le composé testé.

Des exemples d'application de constitutions de telles banques peuvent être fournis par des modèles de culture d'hépatocytes, telle la lignée  
25 HepG2, de cellules épithéliales rénales, telle la lignée HK-2, ou de cellules endothéliales, telle la lignée ECV304, traités par des agents toxiques tel l'éthanol, la camptothécine ou le PMA.

Un exemple de choix peut également être fourni par l'utilisation en cosmétologie de modèles de culture de peau traités ou non par des  
30 agents toxiques ou irritants.

La présente demande a donc également pour objet des

banques de criblage différentiel (entre les ADNc et les ARN), réalisées à partir d'organes, de tissus ou de cultures cellulaires de référence traités avec des composés chimiques représentatifs de grandes classes d'agents toxiques selon les abaques décrites dans la littérature. L'invention concerne

5 aussi l'étalement de ces banques sur des filtres ou sur des supports connus de l'homme de l'art (nitrocellulose, nylon...). Avantageusement ces supports peuvent être des chips ou puces qui définissent ainsi des puces de génotoxicité. L'invention concerne en outre l'exploitation qui peut être faite du séquençage des différents clones qui constituent ces banques dans le but

10 d'élucider les mécanismes mis en jeu par l'action des différents toxiques, ainsi que l'utilisation de ces banques pour les hybrider avec des sondes provenant de cellules ou de tissus traités par un composé chimique ou un produit pharmaceutique dont on veut évaluer la toxicité. Avantageusement, l'invention concerne des banques d'acides nucléiques telles que définies ci-

15 avant, préparées à partir de cellules de la peau traitées en différentes conditions toxiques. L'invention concerne en outre un kit comprenant ces différentes banques différentielles de la peau.

L'invention concerne également l'utilisation des méthodes, acides nucléiques ou banques décrits ci-dessus pour évaluer (prédire) ou

20 améliorer le potentiel thérapeutique de composés tests (génopharmacologie).

Dans cette utilisation, le principe mis en application est très proche de celui décrit précédemment. Des banques différentielles de référence sont établies entre les ADNc et les ARN d'une culture cellulaire ou

25 d'un organe dans une situation contrôle et de leur équivalent mimant un modèle de pathologie. L'efficacité thérapeutique d'un produit peut alors être évaluée en suivant sa capacité à antagoniser les variations qualitatives de l'expression génique qui sont spécifiques du modèle pathologique. Cela est mis en évidence par la modification du profil d'hybridation d'une sonde issue

30 du modèle pathologique sur les banques de références: sans traitement, la sonde n'hybride qu'avec la banque qui contient les signatures spécifiques de

la maladie. Après traitement avec un produit efficace, la sonde bien que provenant du modèle pathologique hybride préférentiellement avec l'autre banque, qui porte les signatures du modèle équivalent sain.

A cet égard, l'invention concerne également un procédé de  
5 détermination ou d'évaluation de l'efficacité thérapeutique d'un composé test sur un échantillon biologique donné comprenant l'hybridation :

- de banques différentielles entre les ADNc et les ARN dudit échantillon biologique à l'état sain et à (différents stades de développement de) l'état pathologique avec,
- 10 - une préparation d'acides nucléiques de l'échantillon biologique traité par ledit composé test, et
- l'évaluation du potentiel thérapeutique du composé test par analyse du degré d'hybridation avec les différentes banques.

Un exemple d'une telle application peut être fourni par un  
15 modèle d'apoptose mimant certains aspects de la neurodégénérescence qui sont antagonisés par des facteurs trophiques de référence. Ainsi les cellules dérivées de phéochromocytômes PC12 différenciées en corps neuronaux en présence de NGF entrent en apoptose par retrait de ce facteur de croissance. Cette apoptose est accompagnée par expression de nombreux  
20 marqueurs de mort cellulaire programmée dont plusieurs sont régulés par épissage alternatif et dont l'apparition est inhibée par action d'IGF1. Deux banques issues de criblage différentiel qualitatif sont établies à partir d'ARNm extraits de cellules PC12 différenciées entrées en apoptose par retrait de NGF d'une part et à partir de PC12 différenciées sauvegardées de  
25 l'apoptose par ajout d'IGF1 d'autre part. Sur ces banques peuvent être hybridées des sondes réalisées à partir d'ARNm extraits de PC12 différenciées entrées en apoptose et dont la survie est améliorée par traitement avec un produit neuroprotecteur à tester. L'efficacité de l'inversion des caractéristiques qualitatives induites par le composé test peut donc être  
30 appréciée par la capacité de la sonde à hybrider spécifiquement les clones spécifiques de la banque représentative des cellules dont la survie est

améliorée. Ce test peut par la suite être utilisé pour tester l'efficacité de dérivés du composé ou de tout autre nouvelle famille de composés neuroprotecteurs et en améliorer le profil pharmacologique.

5 Dans un mode de réalisation particulier, le procédé de l'invention permet d'évaluer l'efficacité d'un composé test neuroprotecteur par hybridation avec une banque différentielle selon l'invention entre une cellule nerveuse saine et cette cellule présentant un modèle de neurodégénérescence.

10 Dans un autre mode, il s'agit de tester un composé anti-tumoral sur des banques différentielles établies à partir d'un échantillon de cellules tumorales et un échantillon sain.

Comme indiqué ci-avant, le procédé de l'invention peut en outre être utilisé pour améliorer les propriétés d'un composé, en testant différents dérivés pour leur capacité à induire un profil d'hybridation proche de la  
15 banque représentative de l'échantillon sain.

L'invention concerne également l'utilisation des méthodes, acides nucléiques ou banques décrits ci-dessus en pharmacogénomique, i.e., pour évaluer (prédire) la réponse d'un patient à un composé ou traitement test.

20 La pharmacogénomique a pour ambition d'établir des profils génétiques de patients afin de déterminer quel traitement est susceptible d'être couronné de succès pour une pathologie donnée. Les techniques décrites dans la présente invention permettent à cet égard d'établir des banques d'ADNc représentatives des différences qualitatives qui existent  
25 entre une situation pathologique qui répond à un traitement donné et une autre qui répond peu ou mal, susceptible d'être l'objet d'une autre stratégie thérapeutique. Ces banques de références établies, elles peuvent être hybridées avec des sondes réalisées à partir d'ARN messagers de patients. Les résultats d'hybridation permettent de savoir quel patient a un profil  
30 d'hybridation correspondant à la situation de répondeur ou de non répondeur et ainsi d'affiner le choix de traitement.



Dans cette application, le but est d'une part de proposer en fonction du patient le traitement le plus approprié, le plus susceptible d'être couronné de succès et d'autre part d'enrôler dans un traitement les patients les plus susceptibles d'y répondre avec succès. Comme dans les autres applications, deux banques de criblage différentiel qualitatif sont réalisées: l'une à partir d'un modèle ou d'un échantillon pathologique connu pour répondre à un traitement, l'autre à partir d'un autre modèle ou échantillon pathologique qui répond peu ou mal à l'action thérapeutique. Ces deux banques sont ensuite hybridées avec des sondes provenant d'ARNm extraits de biopsies de différents patients. Selon que ces sondes hybrident préférentiellement avec les épissages alternatifs spécifiques de l'une ou l'autre situation, les patients peuvent être répartis en répondeurs et en non répondeurs au traitement de référence qui a défini les modèles de départ.

A cet égard, l'invention concerne également un procédé de détermination ou d'évaluation de la réponse d'un patient à un composé ou traitement test comprenant l'hybridation :

- de banques différentielles entre les ADNc et les ARN d'un échantillon biologique répondeur audit composé/traitement et d'un échantillon biologique non-répondeur ou mal-répondeur audit composé/traitement, avec,
- une préparation d'acides nucléiques d'un échantillon biologique pathologique du patient, et
- l'évaluation du potentiel répondeur du patient par analyse du degré d'hybridation avec les différentes banques.

Un exemple de choix de l'apport du criblage différentiel qualitatif à la pharmacogénomique est constitué par un criblage différentiel qualitatif entre deux tumeurs de même origine histologique, l'une régressant lors du traitement par un composé antitumoral (par exemple un transfert d'un ADNc codant pour la protéine p53 sauvage par thérapie génique), l'autre se montrant réfractaire à ce traitement. La première retombée de la constitution de banques de différences qualitatives entre ces deux situations est de

déterminer, par analyse des clones qui constituent ces banques, quels mécanismes moléculaires sont mobilisés lors de la régression du premier modèle et ne sont pas présents dans le deuxième.

Ensuite, l'utilisation de filtres ou tout autre support présentant les  
5 ADNc de ces banques permet de réaliser des hybridations avec des sondes dérivées d'ARNm de biopsies de tumeurs dont on veut anticiper la réponse audit traitement. Ces résultats permettent ainsi de proposer un enrôlement optimisé des patients dans un protocole clinique.

Un exemple particulier de ce procédé consiste à déterminer la  
10 réponse de tumeurs à un traitement par le gène suppresseur de tumeur p53. Il a en effet été décrit que certains patients et certaines tumeurs répondent plus ou moins bien à ce type de traitement (Roth et al., Nature Medicine, 2 (1995) 958). Il est donc important de pouvoir déterminer quels types de tumeurs et/ou quels patients sont sensibles à un traitement par thérapie  
15 génique par p53 sauvage, afin d'optimiser le traitement et de favoriser l'enrôlement des patients dans les essais cliniques en cours. Le procédé de l'invention permet avantageusement de faciliter ces étapes en proposant des banques spécifiques des qualités de cellules répondeuses et de cellules non répondeuses à p53. Des exemples de modèles cellulaires p53-sensibles ou  
20 résistants sont décrits par exemple par Sabbatini et al. (Genes Dev. 9 (1995) 2184) ou par Roemer et al. (Oncogene 12 (1996) 2069). L'hybridation de ces banques avec des sondes dérivées de biopsies de patients permet aisément d'évaluer leur potentiel répondeur. En outre les banques spécifiques permettent également d'identifier des acides nucléiques impliqués dans la  
25 réponse à p53.

La présente demande concerne donc également l'établissement de banques de criblage différentiel à partir d'échantillons pathologiques, ou de modèles de pathologie, qui répondent différemment à au moins un agent pharmacologique. Ces banques peuvent être des  
30 banques restreintes, complexes ou autologues comme définies ci-dessus. Elle concerne aussi l'étalement de ces banques sur des filtres ou sur des

déterminer, par analyse des clones qui constituent ces banques, quels mécanismes moléculaires sont mobilisés lors de la régression du premier modèle et ne sont pas présents dans le deuxième.

Ensuite, l'utilisation de filtres ou tout autre support présentant les  
5 ADNc de ces banques permet de réaliser des hybridations avec des sondes dérivées d'ARNm de biopsies de tumeurs dont on veut anticiper la réponse audit traitement. Ces résultats permettent ainsi de proposer un enrôlement optimisé des patients dans un protocole clinique.

Un exemple particulier de ce procédé consiste à déterminer la  
10 réponse de tumeurs à un traitement par le gène suppresseur de tumeur p53. Il a en effet été décrit que certains patients et certaines tumeurs répondent plus ou moins bien à ce type de traitement (Roth et al., Nature Medicine, 2 (1995) 958). Il est donc important de pouvoir déterminer quels types de tumeurs et/ou quels patients sont sensibles à un traitement par thérapie  
15 génique par p53 sauvage, afin d'optimiser le traitement et de favoriser l'enrôlement des patients dans les essais cliniques en cours. Le procédé de l'invention permet avantageusement de faciliter ces étapes en proposant des banques spécifiques des qualités de cellules répondeuses et de cellules non répondeuses à p53. Des exemples de modèles cellulaires p53-sensibles ou  
20 résistants sont décrits par exemple par Sabbatini et al. (Genes Dev. 9 (1995) 2184) ou par Roemer et al. (Oncogene 12 (1996) 2069). L'hybridation de ces banques avec des sondes dérivées de biopsies de patients permet aisément d'évaluer leur potentiel répondeur. En outre les banques spécifiques permettent également d'identifier des acides nucléiques impliqués dans la  
25 réponse à p53.

La présente demande concerne donc également l'établissement de banques de criblage différentiel à partir d'échantillons pathologiques, ou de modèles de pathologie, qui répondent différemment à au moins un agent pharmacologique. Ces banques peuvent être des  
30 banques restreintes, complexes ou autologues comme définies ci-dessus. Elle concerne aussi l'étalement de ces banques sur des filtres ou sur des

supports connus de l'homme de l'art (nitrocellulose, nylon...).  
Avantageusement ces supports peuvent être des chips ou puces qui  
définissent ainsi des puces de pharmacogénomique. L'invention porte encore  
sur l'exploitation qui peut être faite du séquençage des différents clones qui  
5 constituent ces banques dans le but d'élucider les mécanismes qui président  
aux différences de réponses d'échantillons pathologiques à différents  
traitements, ainsi que l'utilisation de ces banques pour les hybrider avec des  
sondes provenant de biopsies provenant de situations pathologiques dont on  
veut anticiper la réponse au traitement de référence qui définit les banques.

10 La présente invention décrit ainsi que des variations dans les  
formes et/ou profils d'épissages constituent des sources de marqueurs de  
pharmacogénomique, c'est-à-dire des sources de marqueurs permettant la  
mise en évidence de la capacité et de la manière d'un patient à répondre à  
des traitements. A cet égard, l'invention a donc également pour objet  
15 l'utilisation de l'intervariabilité, entre individus, des isoformes générées par  
épissage alternatif (analyse du spliceome) comme source de marqueurs de  
pharmacogénomique. L'invention concerne aussi l'utilisation de modifications  
d'épissage induites par des traitements comme source de marqueurs de  
pharmacogénomique. Ainsi, comme expliqué ci-avant, les methodologies  
20 DATAS de l'invention permettent de générer des acides nucléiques  
représentatifs des différences qualitatives entre deux échantillons  
biologiques. Ces acides nucléiques, ou des formes dérivées (sondes,  
amorces, acides complémentaires, etc.) peuvent être utilisés pour l'analyse  
du spliceome de sujets, en vue de mettre en évidence leur capacité/manière  
25 de répondre à des traitements, ou leur prédisposition à tel  
traitement/pathologie, etc.

Ces différents exemples généraux illustrent l'intérêt des  
banques de criblage différentiel qualitatif dans des études de génotoxicité,  
30 génopharmacologie, pharmacogénomique ainsi que dans des recherches de  
cibles d'intérêt diagnostique ou thérapeutique. Ces banques sont issues du

clonage des différences qualitatives qui existent entre deux situations physiopathologiques. Puisqu'une autre utilisation des ADNc représentatifs de ces différences qualitatives est de constituer des sondes destinées à cribler une banque d'ADN génomique dont les caractéristiques ont été décrites ci avant, une telle approche peut être également mise en œuvre pour toute étude de génotoxicité, génopharmacologie et pharmacogénomique ainsi que d'identification de gène. Dans les études de génotoxicité par exemple, les clones génomiques restreints par la taille de leurs insertions statistiquement à un seul intron ou à un seul exon sont classés sur des filtres en fonction de leur hybridation avec des sondes DATAS provenant de l'analyse différentielle qualitative entre une population cellulaire ou un tissu de référence et les mêmes cellules ou tissu traités par un composé toxique de référence. Ces clones représentatifs des différentes classes de toxicité étant sélectionnés, il peut ensuite être procédé à une hybridation de ces clones avec une sonde dérivée des ARN messagers totaux d'une même population cellulaire ou d'un même tissu traité par un composé dont on veut apprécier le potentiel toxique.

D'autres avantages et applications de la présente invention apparaîtront à la lecture des exemples qui suivent, qui doivent être considérés comme illustratifs et non limitatifs. Les champs d'application de l'invention sont représentés sur la figure 7.

### LEGENDE DES FIGURES

Figure 1. Représentation schématique des criblages différentiels selon l'invention (figure 1A) utilisant une (figure 1B) ou deux (figure 1C) hybridations, et utilisation des acides nucléiques (figure 1D).

Figure 2. Représentation schématique décrivant l'obtention d'hybrides ARN/ADN permettant de caractériser les séquences ARN simple brin, signatures spécifiques de l'état pathologique ou de l'état sain.

Figure 3. Représentation schématique décrivant un moyen permettant d'isoler et de caractériser par séquençage les séquences d'ARN simple brin spécifiques d'une situation pathologique ou d'une situation saine.

5        Figure 4. Représentation schématique décrivant un autre moyen permettant de caractériser par séquençage tout ou une partie des ARNs simple brin spécifiques d'une situation pathologique ou d'une situation saine.

Figure 5. Représentation schématique permettant d'isoler les produits  
10 d'épissages alternatifs grâce à des structures R-loop

Figure 6. Représentation schématique du criblage différentiel qualitatif par restriction de boucles (formation d'homoduplexes ADN<sub>cdb</sub>/ADN<sub>c</sub> et extractions des informations, Figure 6A) et description des informations  
15 obtenues (Figure 6B).

Figure 7. Apports du criblage différentiel qualitatif aux différentes étapes de la recherche et du développement pharmaceutique.

20        Figure 8. Isolation d'un domaine différentiellement épissé dans le modèle grb2/grb33. A) Production des ARNs synthétiques de grb2 et de grb33. B) Suivi des premières étapes de DATAS conduisant à la caractérisation d'un fragment ARN correspondant au domaine différentiellement épissé ; 1 : ARN de grb2, 2 : Hybridation entre l'ARN de grb2 et l'ADN<sub>c</sub> de grb33, 3 : Hybridation entre l'ARN de grb2 et l'ADN<sub>c</sub> de grb2, 4 : Hybridation entre l'ARN de grb2 et de l'eau, 5 : Surnageant après passage sur billes Streptavidine de (2), 6 : Surnageant après passage sur billes Streptavidine de (3), 7 : Surnageant après passage sur billes Streptavidine de (4), 8 : Digestion du duplex ARN grb2 / ADN<sub>c</sub> grb33 à la  
25 Rnase H, 9 : Digestion du duplex ARN grb2 / ADN<sub>c</sub> grb2 à la Rnase H, 10 : Digestion de l'ARN grb2 à la Rnase H, 11 : pareil que (8) après passage sur  
30

colonne d'exclusion, 12 : pareil que (9) après passage sur colonne d'exclusion, 13 : pareil que (10) après passage sur colonne d'exclusion.

Figure 9 . Représentation des populations d'ARNs non appariées  
5 issues de la digestion par la Rnase H à partir de duplexes ARN / ADNc simple brin provenant de cellules HepG2 traitées ou non à l'ethanol.

Figure 10 . Représentation des populations d'ADNc double brin  
générées par une des variantes de DATAS. 1 à 12 : PCRs à partir de  
10 populations de boucles d'ARNs issues de la digestion à la Rnase H, 13 : PCR à partir d'ADNc total.

Figure 11. Application de la variante de DATAS faisant intervenir les  
ADNc double brin sur le modèle grb2/grb33. A) Analyse sur gel d'agarose  
15 des complexes après hybridation : 1 : ADNc double brin grb2 / ARN grb33, 2 : ADNc double brin grb2 /ARN grb2, 3 : ADNc double brin grb2 / eau. B) Digestion des échantillons 1,2 et 3 de A) par la nucléase S1 et la nucléase " Mung Bean " : 1 à 3 : complexes 1 à 3 avant traitement au glyoxal ; 4 à 6 : complexes 1 à 3 après traitement au glyoxal ; 7 à 9 : Digestions de 1 à 3 par  
20 nucléase S1 ; 10 à 12 : Digestions de 1 à 3 par nucléase Mung Bean.

Figure 12. Application de la variante de DATAS faisant intervenir les  
ADNc simple brin et la Rnase H sur un système de cellules HepG2 traitées  
ou non à l'ethanol 0,1M pendant 18 heures. Les inserts clonés ont été  
25 transférés sur membrane après électrophorèse sur gel d'agarose et soumis à hybridation à l'aide de sondes correspondant aux situations traitées (Tr) ou non (NT).

Figure 13. Mode opératoire pour évaluer le potentiel toxique d'un  
30 produit.

Figure 14. Mode opératoire pour suivre l'efficacité d'un produit.

Figure 15. Mode opératoire pour étudier la susceptibilité d'une situation pathologique à un traitement.

5

Figure 16. Analyse d'hybridation différentielle de clones issus de DATAS à partir d'ARNs extraits de cellules induites et d'ADNc extraits de cellules non induites. A) Utilisation de colonies bactériennes déposées et lysées sur membrane. B) Southern Blot effectué à partir d'une sélection de clones de A.

10

Figure 17. Séquences nucléotidique et peptidique de  $\Delta$ SHC (SEQ ID NO: 9 et 10).

15

Figure 18. Tests de cytotoxicité et d'apoptose sur cellules HepG2 traitées A) à l'éthanol ; B) à la camptothécine ; C) au PMA.

Figure 19. Réactions de RT-PCR effectuées à partir d'ARNs extraits de cellules HepG2 traitées ou non (N.T) par l'éthanol (Eth.), la camptothécine (Camp.) et le PMA (PMA) permettant l'amplification de fragments correspondants à des domaines de MACH-a, BCL-X, FASR et beta-actine comme contrôle de normalisation.

20

Dans les exemples et la description de l'invention, il est fait références aux séquences de la Liste de Sequences, qui contient le texte libre suivant:

25

<223> OLIGO

<223> OLIGO

<223> OLIGO

<223> OLIGO

30

<223> OLIGO

<223> OLIGO



<223> OLIGO  
<223> OLIGO  
<223> OLIGO  
<223> OLIGO  
5 <223> OLIGO  
<223> OLIGO

### EXEMPLES

#### 10 1. CLONAGE DIFFÉRENTIEL DES EPISSAGES ALTERNATIFS ET AUTRES MODIFICATIONS QUALITATIVES DES ARNS EN UTILISANT DES ADNc SIMPLE-BRINS

Les ARN messagers correspondant à deux situations, l'une normale  
15 (mN) et l'autre pathologique (mP), sont isolés à partir de biopsies ou de  
cellules en culture. Ces ARN messagers sont convertis en ADN  
complémentaires (cN) et (cP) à l'aide de reverse transcriptase (RT). Des  
hybrides mN/cP et cN/mP sont ensuite réalisés en phase liquide (se reporter  
au schéma de la figure 2 illustrant un des deux cas aboutissant à la  
20 formation de cN/mP).

Ces hybrides sont avantageusement réalisés en émulsion phénolique  
(technique PERT ou Phenol Emulsion DNA Reassociation Technique)  
maintenue par thermocycles (Miller, R.,D. and Riblet, R., 1995, Nucleic Acids  
Research, vol 23, n°12, pp 2339-2340). Typiquement, cette étape  
25 d'hybridation est réalisée entre 0,1 à 1 µg d'ARN polyA+ et 0,1 à 2µg d'ADN  
complémentaire dans une émulsion formée d'une phase aqueuse (tampon  
phosphate de sodium 120mM, NaCl 2,5M, EDTA 10mM) et d'une phase  
organique représentant 8% de la phase aqueuse et constituée de phénol  
bidistillé.

30 Une autre technique est également avantageusement  
employée de façon à obtenir des hétéroduplex : à l'issue de la transcription

inverse, l'ADNc néosynthétisé est séparé de l'amorce oligodT biotinylée sur colonne d'exclusion. 0,1 à 2µg de cet ADNc est coprécipité avec 0,1 à 1µg d'ARN polyA+ en présence de 0,3M d'acétate de sodium et de deux volumes d'éthanol. Ces acides nucléiques coprécipités sont repris dans 30µl d'un  
5 tampon d'hybridation qui contient 80% de formamide, 40mM de PIPES (piperazinebis(2-ethanesulfonic acid)) pH6,4, 0,4M de NaCl et 1mM d'EDTA.

Les acides nucléiques en solution sont dénaturés par chauffage 10mn à 85°C puis leur hybridation est réalisée pendant au moins 16h et jusqu'à 48h à 40°C.

10 L'intérêt de la technique d'hybridation en formamide est de permettre des conditions de plus forte sélectivité lors de l'appariement des brins d'ADNc et d'ARN.

A l'issue de chacune de ces deux techniques d'hybridation, nous disposons d'hétéroduplex ARN/ADN dont la perfection d'appariement dépend de l'efficacité de la RT à synthétiser la longueur totale des ADNc.  
15 Demeurent également sous forme de simples brins les régions d'ARN (et d'ADN) qui correspondent aux épissages alternatifs qui différencient les deux états physiopathologiques étudiés.

Le but de la méthode est ensuite de caractériser l'information  
20 génétique portée par ces boucles d'épissage.

Pour cela, les hétéroduplex sont purifiés par capture des ADNc (amorçés avec des oligodT biotinylés) grâce à des billes portant des groupements streptavidines. Avantagusement ces billes sont de billes douées de propriétés magnétiques, ce qui permet de les séparer des ARN  
25 non engagés dans les hétéroduplex par action d'un séparateur magnétique. De telles billes et de tels séparateurs sont disponibles commercialement.

Sont isolés à ce stade de la procédure les hétéroduplex et les ADNc non engagés dans des hybridations avec des ARN. Ce matériel est ensuite soumis à l'action de la RNaseH qui va spécifiquement hydrolyser les  
30 régions d'ARN hybridées avec les ADNc. Les produits résultant de cette hydrolyse sont d'une part les ADNc et d'autre part les fragments d'ARN qui

correspondent aux boucles d'épissage ou aux régions non hybridées du fait du rendement partiel de la transcriptase inverse. Les fragments d'ARN sont séparés de l'ADN par séparation magnétique selon le même mode opératoire que celui mentionné plus haut et par digestion avec de la DNase  
5 exempte de toute contamination par une activité Rnase.

### 1.1. Validation de la technique DATAS sur les variants d'épissage du gène Grb2

Une mise en évidence de la faisabilité de cette approche a été  
10 réalisée sur un système in-vitro utilisant un ARN correspondant à la région codante de Grb2 d'une part et un ADNc simple brin complémentaire à la région codante de Grb3.3. Grb2 est un gène possédant une phase codante de 651 paires de bases. Grb33 est une isoforme de grb2 générée par épissage alternatif et comprenant une délétion de 121 paires de bases dans  
15 le domaine fonctionnel SH2 de grb2 (Fath et.al, Science (1994), 264, 971-4). Les ARNs de Grb2 et de Grb33 sont synthétisés selon les techniques connues de l'homme du métier à partir d'un plasmide contenant la séquence codante de Grb2 ou de Grb33 sous contrôle du promoteur T7 à l'aide du kit RiboMax (Promega). L'analyse des produits démontre une synthèse  
20 homogène (figure 8A). Dans un but de visualisation, l'ARN de Grb2 a également été rendu radioactif par incorporation d'une base marquée lors de la transcription in-vitro à l'aide du kit RiboProbe (Promega). Les ADNc de Grb2 et de Grb3.3 ont été synthétisés par transcription inverse à partir des ARNs synthétiques produits ci-dessus, du kit Superscript II (Life  
25 Technologies) et d'une amorce oligonucleotidique biotinylée commune à Grb2 et à Grb33 correspondant au complémentaire de la séquence (618-639) de Grb2. Les ARNs et ADNcs ont été traités selon les indications des fournisseurs (promega, Life Technologies), purifiés sur colonne d'exclusion (Rnase free sephadex G25 ou G50, 5 Prime, 3 Prime) et quantifiés par  
30 spectrophotométrie.

Les premières étapes de DATAS ont été appliquées en

associant, en suspension 10ng d'ARN marqué de Grb2 avec :

1. 100 ng d'ADNc de grb33 biotinylé,
2. 100 ng d'ADNc de grb2 biotinylé,
3. de l'eau

5 dans 30 µl d'un tampon d'hybridation qui contient 80% de formamide, 40 mM PIPES (pH 6,4), 0,4 M NaCl, 1 mM EDTA. Les acides nucléiques sont dénaturés par chauffage 10 mn à 85 °C, puis l'hybridation est réalisée pendant 16 heures à 40°C. Après capture à l'aide de billes streptavidine, les échantillons sont traités à la RNase H comme décrit précédemment.

10 L'analyse de ces étapes est réalisée par électrophorèse sur gel d'acrylamide à 6% suivi d'un traitement des gels à l'aide d'un Instant Imager (Packard Instruments) permettant la qualification et la quantification des espèces issues de l'ARN de grb2 marqué (figure 8B). Ainsi, les puits 2,3 et 4 indiquent que les duplexes grb2/grb33 et grb2/grb2 se sont formés de façon  
15 quantitative. La migration du complexe grb2/grb33 est retardée par rapport à celle de l'ARN de grb2 (puits 2) alors que celle du complexe grb2/grb2 est augmentée (puits 3). Les puits 5,6 et 7 correspondent aux échantillons non retenus par les billes streptavidine démontrant que 80% des complexes grb2/grb33 et grb2/grb2 ont été retenus sur les billes alors que l'ARN de grb2  
20 seul, non biotinylé, se retrouve exclusivement dans le surnageant des billes. Le traitement à la Rnase H libère, outre les nucléotides libres qui migrent plus vite que le bleu de Bromophénol (BPB) une espèce migrant en deçà du bleu de xylène Cyanol (XC) (marqué par une flèche sur la figure) et ce spécifiquement dans le puits 8 correspondant au complexe grb2/grb33 par  
25 rapport aux puits 9 et 10 qui correspondent au complexe grb2/grb2 et à l'ARN de grb2. Les puits 11,12 et 13 correspondent aux puits 8,9 et 10 après passage des échantillons sur une colonne d'exclusion pour éliminer les nucléotides libres. La migration observée dans les puits 8 et 11 est celle attendue pour une molécule d'ARN correspondant à la délétion de 121  
30 nucléotides différenciant grb2 de grb33.

Ce résultat montre bien la possibilité d'obtenir les boucles

d'ARN générées par la formation d'hétéroduplex entre deux séquences dérivées de deux isoformes d'épissage.

5     1.2. Application de la technique DATAS à la génération de banques qualitatives de cellules hépatiques dans un état sain et toxique

Une situation plus complexe a été étudiée. Dans le cadre de l'application de la technologie DATAS comme outil prédictif de toxicité de molécules, des cellules humaines de type hépatocytaire, HepG2, ont été traitées par de l'ethanol 0,1 M pendant 18 heures. Les ARNs ont été extraits  
10     à partir des cellules traitées ou non. La variante de DATAS décrite ci-dessus (préparation d'ADNc sb biotinylés, hybridations croisées en phase liquide, application d'un champs magnétique pour séparer les espèces, traitement RNaseH) a été appliquée avec les cellules non traitées en situation de référence (ou situation A) et les cellules traitées en situation test (ou situation  
15     B) (figure 9). Les ARNs extraits n'étant pas marqués radioactivement, la visualisation des populations d'ARN générées par digestion à la RnaseH est réalisée en effectuant une réaction d'échange du phosphate en 5' des ARNs par un phosphate marqué, à l'aide de polynucleotide kinase de T4 et de gamma-P<sup>32</sup>ATP. Ces marquages sont ensuite déposés sur un gel  
20     d'acrylamide/urée et analysés par exposition à l'aide d'un Instant Imager (Packard Instruments). Des signatures complexes issues des hybridations A/B et B/A peuvent alors être visualisées avec un premier groupe de signaux migrant faiblement dans le gel et correspondant à des séquences d'acides nucléiques de taille importante et un deuxième groupe de signaux migrant  
25     entre 25 et 500 nucleotides. Ces signatures sont d'intensité beaucoup plus faible à partir de la situation A/A suggérant que l'ethanol peut induire une reprogrammation de l'épissage des ARNs, traduite par l'existence de signaux en A/B et B/A.

30     1.3. Clonage et Préparation de banques à partir des acides nucléiques identifiés.

Plusieurs variantes expérimentales sont ensuite envisageables pour cloner ces fragments d'ARN résistant à l'action de la Rnase H:

5 A. Une première approche consiste à isoler ces boucles et à les cloner (Figure 3).

Selon cette approche, il est procédé à une ligation d'oligonucléotides à chacune des extrémités par action de la RNA ligase selon les conditions connues de l'homme de l'art. Ces oligonucléotides sont ensuite utilisés comme amorces pour effectuer une RT PCR. Les produits de  
10 PCR sont clonés et criblés avec des sondes d'ADN complémentaires totales correspondant aux deux situations physiopathologiques d'intérêt. Seuls les clones hybridant préférentiellement avec une seule des deux sondes contiennent les boucles d'épissage qui sont ensuite séquencées et/ou utilisées pour générer des banques.

15

B. La seconde approche (Figure 4) consiste à effectuer une transcription inverse sur l'ARN simple brin libéré des hétéroduplex après action de la RNaseH, initiée à l'aide d'amorces au moins en partie aléatoires. Ainsi, il peut s'agir d'amorces aléatoires en 3' et en 5', d'amorces aléatoires  
20 en 3' et déterminées en 5', ou encore d'oligonucléotides semi-aléatoires, c'est-à-dire comprenant une zone de dégénérescence et une zone définie.

Selon cette stratégie, les amorces sont donc susceptibles de s'hybrider soit n'importe où sur l'ARN simple brin, soit à chaque succession de bases fixée par le choix de l'amorce semi-aléatoire. Une PCR avec des  
25 amorces correspondant aux oligonucléotides décrits ci-dessus permet ensuite d'obtenir des séquences dérivées des boucles d'épissage.

La figure 10 (puits 1 à 12) montre l'analyse sur gel d'acrylamide des fragments de PCR obtenus à partir de plusieurs essais DATAS et couplée à l'utilisation des oligonucléotides semi-aléatoires suivants:

30 GAGAAGCGTTATNNNNNNNAGGT (SEQ ID NO: 1, X=T)

GAGAAGCGTTATNNNNNNNAGGA (SEQ ID NO: 1, X=A)

GAGAAGCGTTATNNNNNNNAGGC (SEQ ID NO: 1, X=C)

GAGAAGCGTTATNNNNNNNAGGG (SEQ ID NO: 1, X=G)

La comparaison avec la complexité des signaux obtenus en  
5 utilisant les mêmes oligonucléotides, mais de l'ADNc total comme matrice  
(puits 13) démontre que DATAS a permis de filtrer ("profiler") des  
informations correspondant à des différences qualitatives.

Cette variante a été utilisée afin de cloner un événement  
correspondant au domaine ARN de grb2 généré par action de la Rnase H à  
10 partir du duplex ARN grb2 / ADNc simple brin de grb33 selon le protocole  
décrit précédemment (exemple 1.1.). A cette fin, un oligonucléotide de  
séquence : GAGAAGCGTTATNNNNNNNTCCC (SEQ ID NO: 2), choisi sur  
le modèle GAGAAGCGTTATNNNNNNNWXYZ (dans lequel N est défini  
comme précédemment, W, X et Y représentent chacun une base fixe  
15 déterminée, et Z représente soit une base déterminée soit un groupe 3'-OH,  
SEQ ID NO: 3) et sélectionné pour amplifier un fragment dans la délétion de  
grb2 a été utilisé, permettant de générer un fragment PCR dont le clonage et  
le séquençage a démontré qu'il était effectivement issu du domaine délété  
de grb2 (194-281 dans grb2).

20

Ces deux approches permettent donc la production de  
compositions d'acides nucléiques représentatifs des épissages différentiels  
dans les deux situations testées qui peuvent être employées comme sondes  
ou pour construire des banques d'ADNc de différences qualitatives. La  
25 capacité de la technologie DATAS à générer des banques profilées d'ADNc  
représentatives de différences qualitatives est également illustrée par  
l'exemple 1.4. suivant.

#### 1.4. Production de banques profilées représentatives de 30 cellules endothéliales humaines

Cet exemple a été réalisé à partir d'une lignée de cellules

endothéliales humaines (ECV304). L'analyse qualitative de l'expression génétique a été réalisée à partir d'ARN cytosoliques extraits de cellules en prolifération, d'une part, et de cellules en anoïkis (apoptose par privation de support d'attachement), d'autre part.

5 Les cellules ECV ont été cultivées en milieu 199 supplémenté en sels de Earle (Life Sciences). Leur mise en anoïkis a été réalisée par passage pendant 4 heures sur boîtes de culture traitées au polyHEMA. Lors de la préparation des ARN, les cellules ont été lysées dans un tampon contenant du Nonidet P-40. Les noyaux sont ensuite écartés par  
10 centrifugation. La solution d'extrait cytoplasmique a été ensuite ajustée de manière à fixer de façon spécifique l'ARN à la matrice de silice Rneasy selon les instructions de la société Qiagen. Après lavage, les ARN totaux sont élués dans de l'eau traitée au DEPC. Les ARNs messagers sont préparés à partir des ARNs totaux par séparation sur billes magnétiques Dynabeads  
15 oligo (dT)<sub>25</sub> (Dyna). Après avoir mis en suspension les billes dans un tampon de fixation, l'ARN total est incubé pendant 5 min à température ambiante. Après séparation magnétique et lavage, les billes sont reprises dans un tampon d'élution pour une incubation à 65°C qui libère les ARNs messagers.

20 Les synthèses d'ADN premier brin sont effectuées à partir des ARNs messagers en utilisant la Reverse Transcriptase SuperScript II ou ThermoScript (Life Technologies) à l'aide d'amorces olido (dT). Après RnaseH, les nucléotides libres sont éliminés par passage sur colonne Séphadex G50 (5 Prime- 3 Prime). Après extraction au phénol /  
25 Chloroforme et précipitation à l'éthanol, les échantillons sont quantifiés par absorbance UV.

Les quantités requises d'ARN et d'ADNc (en l'occurrence 200ng de chaque) sont combinées et précipitées à l'éthanol. Les échantillons sont repris dans un volume de 30µl dans un tampon  
30 d'hybridation (Hepes (pH 7.2) 40 mM, NaCl 400mM, EDTA 1mM) supplémenté de formamide désionisée (80% (v/v), sauf indication contraire).



Après dénaturation 5 min à 70°C, les échantillons sont incubés sur la nuit à 40°C.

Les billes Streptavidine (Dyna) sont lavées puis reconditionnées dans un tampon de fixation (2X= Tris-HCl (pH 7,5) 10 mM, NaCl 2M, EDTA 1 mM). Les échantillons d'hybridation sont amenés à un volume de 200 µl avec de l'eau puis ajustés à 200 µl de billes pour une incubation de 60 min à 30°C. Après capture sur aimant et lavages des billes, celles-ci sont reprises dans 150 µl de tampon RnaseH puis incubées pendant 20 min à 37°C. Après capture sur aimant, les régions non hybridées ont été relarguées dans le surnageant qui est traité à la Dnase puis extrait au phénol acide / chloroforme puis précipité à l'éthanol. Les précipitations à l'éthanol de faibles quantités d'acides nucléiques sont effectuées à l'aide d'un polymère commercial SeeDNA (Amersham Pharmacia Biotech) permettant de récupérer de façon quantitative des acides nucléiques à partir de solutions très diluées (de l'ordre du ng/ml).

La synthèse d'ADNc à partir des échantillons d'ARNs provenant de l'action de la RnaseH est effectuée à partir d'hexanucléotides aléatoires à l'aide de Superscript II Reverse Transcriptase. L'ARN est ensuite dégradé à l'aide d'un mélange de RnaseH et de Rnase T1. L'amorce, les nucléotides non incorporés et les enzymes sont séparés de l'ADNc à l'aide d'une cartouche "GlassMAX Spin". L'ADNc correspondant aux boucles d'épissage est ensuite soumis à une réaction de PCR en utilisant des oligonucléotides de type semi-aléatoire déjà décrits plus haut dans l'invention. En l'occurrence les oligonucléotides choisis sont :

GAGAAGCGTTATNNNNNCCA (SEQ ID NO: 4)

La réaction de PCR est réalisée à l'aide de Taq Polymérase sur 30 cycles :

- Dénaturation initiale : 94°C pdt 1 min.
- 94°C pdt 30 s
- 55°C pdt 30s
- 72°C pdt 30s

- Extension finale : 72°C pdt 5 min.

Les produits de PCR ont été clonés dans le vecteur pGEM-T (Proméga) possédant un T flottant aux extrémités 3' afin de faciliter le clonage de fragments issus de l'activité de la Taq Polymérase. Après transformation dans les bactéries JM109 compétentes (Proméga), les colonies obtenues sont repiquées sur filtre de nitrocellulose, et hybridées avec des sondes dérivées de produits de PCR effectuées sur des ADNc totaux des cellules en prolifération d'une part et en anoïkis d'autre part. Pour ces PCR les mêmes oligonucléotides GAGAAGCGTTATNNNNNCCA sont utilisés. Dans une première réalisation expérimentale, 34 clones hybridant préférentiellement avec la sonde des cellules en apoptose et 13 clones hybridant préférentiellement avec la sonde des cellules en prolifération ont été isolés.

Parmi ces 13 clones, 3 clones contiennent le même fragment d'ADNc qui dérive du domaine SH2 de la protéine SHC.

La séquence de ce fragment est la suivante :

CCACACCTGGCCAGTATGTGCTCACTGGCTTGCAGAGTGGGCAG  
CCAGCCTAAGCATTTCGACTGG (SEQ ID NO: 5)

L'utilisation d'amorces de PCR qui encadrent le domaine SH2 de SHC (oligo5' : GGGACCTGTTTGACATGAAGCCC (SEQ ID NO:6) ; oligo3' : CAGTTTCCGCTCCACAGGTTGC (SEQ ID NO:7)) a permis de caractériser la délétion du domaine SH2 de SHC qui est observée spécifiquement dans les cellules ECV en anoïkis. Avec ce couple d'amorce, un seul produit d'amplification correspondant à un fragment d'ADNc de 382 paires de bases qui contient le domaine SH2 intègre est obtenu à partir d'ARN de cellules ECV en phase exponentielle. Un fragment additionnel de 287 paires de bases est observé lorsque la PCR est réalisée à partir d'ARN de cellules en anoïkis. Ce fragment supplémentaire dérive d'un ARN messager dérivé du messenger de SHC mais présentant une délétion.

La séquence de cette délétion est la suivante :

GTACGGGAGAGCACGACCACACCTGGCCAGTATGTGCTCACTGG

CTTGCAGAGTGGGCAGCCTAAGCATTTGCTACTGGTGGACCCTGAGGG  
TGTG (SEQ ID NO: 8).

Cette délétion correspond aux bases 1198 à 1293 de la phase ouverte du messenger codant pour les formes de 52kDa et 46kDa de la protéine SHC (Pelicci, G. et al, 1992, Cell, 70, pp93-104).

Les données structurales des domaines SH2 ainsi que la littérature indiquent qu'une telle délétion aboutit à la perte de l'affinité pour les phosphotyrosines puisqu'elle englobe les acides aminés impliqués dans les interactions avec les tyrosine phosphorylées (Waksman, G. et al, 1992, Nature, vol358, pp646-653). Les protéines SHC étant des adaptateurs qui connectent différents partenaires par leurs domaines SH2 et PTB (PhosphoTyrosine Binding domain), cette délétion génère donc un dominant négatif naturel de SHC que nous appelons  $\Delta$ SHC. Les domaines SH2 des protéines dont les gènes sont séquencés étant portés par deux exons, il est vraisemblable que la délétion identifiée par la méthode DATAS correspond à un exon alternatif du gène SHC.

Les séquences protéique et nucléique de  $\Delta$ SHC sont les représentées sur la Figure 17 (SEQ ID NO: 9 et 10).

Le domaine SH2 de SHC étant impliqué dans la transduction de nombreux signaux impliqués dans la prolifération et la viabilité cellulaires, l'examen de la séquence de  $\Delta$ SHC permet d'anticiper ses propriétés de dominant négatif sur la protéine SHC et sa capacité d'interférer avec différent signaux cellulaires.

L'invention concerne également cette nouvelle forme épissée de SHC, le domaine protéique correspondant à l'épissage, tout anticorps ou sonde nucléique permettant sa détection dans un échantillon biologique, et leurs utilisation diagnostique ou thérapeutique, par exemple.

L'invention concerne en particulier tout variant de SHC comprenant au moins une délétion correspondant aux bases 1198 à 1293, plus particulièrement une délétion de la séquence SEQ ID NO: 8. L'invention

concerne plus spécifiquement le variant  $\Delta$ SHC ayant la séquence SEQ ID NO: 9, codé par la séquence SEQ ID NO: 10.

L'invention concerne aussi toute sonde nucléique, oligonucléotide ou anticorps permettant d'identifier le variant  $\Delta$ SHC ci-dessus, et/ou toute  
5 altération du rapport SHC/ $\Delta$ SHC dans un échantillon biologique. Il peut s'agir notamment d'une sonde ou oligonucléotide complémentaire de tout ou partie de la séquence SEQ ID NO: 8, ou d'un anticorps dirigé contre le domaine protéique codé par cette séquence. De telles sondes, oligonucléotides ou anticorps permettent de détecter la présence de la forme non épissée (e.g.,  
10 SHC) dans un échantillon biologique.

Les matériels peuvent en outre être utilisés en parallèle avec des sondes, oligonucléotides et/ou anticorps spécifiques de la forme épissée (e.g.,  $\Delta$ SHC), c'est-à-dire correspondant par exemple à la région de jonction résultant de l'épissage (localisée autour du nucléotide 1198 de la séquence  
15 SEQ ID NO: 10).

De tels matériels peuvent être utilisés pour le diagnostic de pathologies liées à une immunodépression (cancer, traitement immunosuppresseur, SIDA, etc.).

L'invention concerne aussi tout procédé de criblage de molécules  
20 basé sur le blocage (i) du domaine épissé dans la protéine SHC (notamment pour induire un état de tolérance immunitaire par exemple dans les maladies autoimmunes ou les rejets de greffes et les cancers) ou (ii) des gains de fonction acquis par la protéine  $\Delta$ SHC.

L'invention concerne en outre l'utilisation thérapeutique de  $\Delta$ SHC, et  
25 notamment pour le traitement de cellules cancéreuses ou de cancers (ex vivo ou in vivo) dans lesquels une hyperphosphorylation de la protéine SHC peut être mise en évidence, par exemple. A cet égard, l'invention concerne aussi tout vecteur, notamment viral, comprenant une séquence codant pour  $\Delta$ SHC. Il s'agit préférentiellement d'un vecteur capable de transfecter des  
30 cellules cancéreuses ou en prolifération, telles que des cellules musculaires

lisses, des cellules endothéliales (resténose), des fibroblastes (fibroses), de préférence d'origine mammifère, notamment humaine. Comme vecteur viral, on peut citer notamment des vecteurs adénoviraux, rétroviraux, AAV, herpès, etc.

5           2. CLONAGE DIFFÉRENTIEL DES EPISSAGES ALTERNATIFS ET AUTRES MODIFICATIONS QUALITATIVES DES ARNS EN UTILISANT DES ADNc DOUBLE-BRINS (FIGURE 5).

Les ARNs messagers correspondants aux situations normales  
10 (mN) et pathologiques (mP) sont produits, ainsi que les ADNs complémentaires double brin correspondants (dsN et dsP) par des protocoles classiques de biologie moléculaire. Des structures de type "R-loop" sont alors obtenues en hybridant mN avec dsP et mP avec dsN dans une solution contenant 70% de formamide. Les domaines nucléiques  
15 différentiellement épissés entre la situation N et P resteront sous forme d'ADN double brin. Les simples brins d'ADN déplacés sont alors traités au glyoxal afin d'éviter le redéplacement du brin d'ARN lors du retrait de la formamide. Après retrait de la formamide et du glyoxal puis traitement à la RNaseH, nous nous retrouvons avec des structures de type abeille, les  
20 ADNs simples brin non appariés représentant les ailes de l'abeille et le domaine double brin apparié d'intérêt représentant le corps de l'abeille. L'utilisation d'enzymes qui dégradent spécifiquement l'ADN simple brin comme la nucléase S1 ou la Mung Bean nucléase permet l'isolation de l'ADN resté sous forme double brin qui est ensuite cloné puis séquencé. Cette  
25 deuxième technique permet l'obtention directe d'une empreinte ADN double brin du domaine d'intérêt comparativement au premier protocole qui produit une empreinte ARN de ce domaine.

Cette approche a été réalisée sur le modèle grb2/grb33 décrit précédemment. L'ADN double brin de grb2 a été produit par amplification  
30 PCR à partir de l'ADNc simple brin de grb2 et de deux amorces nucléotidiques correspondant à la séquence (1-22) de grb2 et à la séquence

complémentaire de (618-639) de grb2. Ce fragment PCR a été purifié sur gel d'agarose, nettoyé sur colonne d'affinité (JetQuick, Genomed) et quantifié par spectrophotométrie. Dans le même temps, deux ARNs synthétiques correspondant aux phases de lecture de grb2 et de grb33 ont  
5 été produits à partir de vecteurs plasmidiques comportant les cDNAs de grb2 ou de grb33 sous contrôle du promoteur T7, à l'aide du kit RiboMax (Promega). Les ARNs ont été purifiés selon les instructions du fournisseur et nettoyés sur colonne d'exclusion (Sephadex G50, 5 prime-3 prime). 600 ng de l'ADN double brin de grb2 (1-639) ont été associés avec :

- 10 1. 3 µg d'ARN de grb33
2. 3 µg d'ARN de grb2
3. de l'eau

dans trois réactions différentes, dans le tampon suivant :

100 mM PIPES (pH 7,2), 35 mM NaCl, 10 mM EDTA, 70% formamide  
15 déionisé (Sigma)

Les échantillons ont été amenés à 56 °C puis refroidis à 44 °C par incrément de -0,2 °C toutes les 10 minutes. Ils sont ensuite conservés à 4 °C. L'analyse sur gel d'agarose révèle des modifications de migrations dans les puits 1 et 2 par rapport au puits 3 contrôle (Figure 11A) indiquant la  
20 formation de nouveaux complexes. Les échantillons sont ensuite traités au glyoxal déionisé (Sigma) (5% v/v ou 1M) pendant 2 h à 12 °C. Les complexes sont ensuite précipités à l'éthanol (0,1M NaCl, 2 volumes d'éthanol), lavés à l'éthanol 70%, séchés puis repris dans de l'eau. Ils sont enfin traités par la RnaseH (Life Technologies), puis par une enzyme capable  
25 de dégrader spécifiquement l'ADN simple brin. La nuclease S1 et la nucléase "Mung Bean" présentent cette propriété et sont disponibles commercialement (Life Technologies, Amersham). De telles digestions (incubations de 5 minutes dans les tampons fournis avec les enzymes) ont été analysées sur gel d'agarose (figure 11B). Des digestions significatives  
30 sont uniquement obtenues à partir des complexes issus de la réaction 1 (grb2/grb33) (figure 11B, puits 7 et 10). Ces digestions semblent plus

complètes avec la nucléase S1 (puits 7) qu'avec la nucléase " Mung Bean " (puits 10). Ainsi, la bande correspondant à une taille légèrement supérieure à 100 paires de bases (indiquée par une flèche dans le puits 7) a été purifiée, clonée dans le vecteur pMos-Blue (Amersham), puis séquencée. Ce  
5 fragment correspond au domaine de 120 paires de bases de grb2, délété dans grb33.

Cette approche peut maintenant être effectuée à partir d'une population totale d'ARN messager et d'une population totale d'ADNc double brin produite selon les techniques connues de l'homme de métier. La  
10 population d'ARN de la situation de référence est hybridée à la population d'ADNc double brin de la situation test et réciproquement. Après application du protocole décrit ci-dessus, les digestions sont déposées sur gel d'agarose afin d'isoler et de purifier les bandes correspondant à des tailles variant entre 50 et 300 paires de bases. Ces bandes sont ensuite clonées dans un  
15 vecteur (pMos-Blue, Amersham) pour donner lieu à une banque d'inserts enrichis en des événements de différences qualitatives.

### 3. CONSTRUCTION DE BANQUES ISSUES DE CRIBLAGES DIFFÉRENTIELS QUALITATIFS.

20

Les deux exemples décrits ci-dessus aboutissent aux clonages d'ADNc représentatifs de toute ou partie des séquences épissées différenciellement entre deux situations physiopathologiques données. Ces ADNc permettent la constitution de banques par insertion de ces ADNc dans  
25 des vecteurs plasmidiques ou phagiques. Ces banques peuvent être présentées sur des filtres de nitrocellulose ou tout autre support connu de l'homme de l'art, tels des chips ou biopuces ou membranes. Ces banques peuvent être conservées au froid, à l'abri de la lumière. Ces banques, une fois déposées et fixées sur support par les techniques classiques, peuvent  
30 être traitées par des composés pour éliminer les bactéries hôtes qui permettent la production des plasmides ou des phages. Ces banques

peuvent également avantageusement être constituées de fragments d'ADNc correspondant aux ADNc clonés mais préparés par PCR de façon à ne déposer sur filtre que les séquences dérivées des événements d'épissages alternatifs.

5 L'une des caractéristiques et en même temps l'une des originalités du criblage différentiel qualitatif est que cette technique aboutit de façon avantageuse non pas à une mais à deux banques différentielles ("paire de banque") qui représentent l'ensemble des différences qualitatives qui existent entre deux situations données. En particulier, l'une des banques  
10 d'épissage différentiel de l'invention représente la signature des qualités de la situation physiologique test par rapport à la situation physiologique de référence, et l'autre banque représente la signature des qualités de la situation physiologique de référence par rapport à la situation physiologique test. Ce couple de banques est également désigné paire de banques ou  
15 "banque différentielle d'épissage".

L'un des apports du criblage différentiel qualitatif étant de permettre d'évaluer le potentiel toxique d'un composé, comme cela est indiqué dans le chapitre suivant, un bon exemple de mise en œuvre de la technologie est l'obtention par DATAS de clones d'ADNc correspondant à  
20 des séquences spécifiques de cellules HepG2 naïves, d'une part, et traitées par de l'éthanol, d'autre part. Ces cellules présentent des signes de cytotoxicité et une dégradation de leur ADN par fragmentation internucléosomale à partir de 18h en présence de 1M d'éthanol. De façon à obtenir des marqueurs précoces de la toxicité éthanolique, les ARN  
25 messagers ont été préparés à partir de cellules naïves et de cellules traitées pendant 18h par de l'éthanol à la concentration de 0,1M. Après mise en œuvre de la variante de DATAS qui utilise l'ADNc simple brin et la Rnase H, les ADNc clonés obtenus ont été amplifiés par PCR, soumis à une électrophorèse sur gel d'agarose et ensuite transférés sur un filtre de nylon  
30 selon les techniques connues de l'homme de l'art. Pour chaque ensemble de clones spécifiques d'une part des différences qualitatives spécifiques de



l'état naïf et d'autre part des séquences spécifiques des cellules traitées par l'éthanol, deux répliques identiques de filtres sont effectuées. Ainsi les empreintes de chaque ensemble de clones sont hybridées d'une part avec une sonde spécifique des cellules non traitées et d'autre part avec une sonde spécifique des cellules traitées par 0,1M d'éthanol pendant 18h.

Le profil d'hybridation différentiel obtenu et présenté sur la figure 12 permet d'apprécier la qualité de la soustraction effectuée lors de la mise en œuvre de la technique DATAS. Ainsi les clones issus de l'hybridation de l'ARNm de cellules non traitées (NT) avec l'ADNc de cellules traitées (Tr) et qui doivent correspondre à des différences qualitatives spécifiques de la situation naïve hybrident préférentiellement avec une sonde représentant la population totale des ARN messagers des cellules non traitées. Réciproquement, les clones issus des produits résistant à la RNase H ayant agi sur les hétéroduplex ARN(Tr)/ADNc(NT) hybrident préférentiellement avec une sonde dérivée de la population totale des ARN messagers des cellules traitées.

Les deux ensembles de clones spécifiques d'une part de la situation traitée et d'autre part de la situation non traitée représentent un exemple de banques de différences qualitatives caractéristiques de deux états cellulaires distincts.

#### 4. UTILISATIONS ET APPORTS DES BANQUES DIFFÉRENTIELLES QUALITATIVES.

Les possibilités d'utilisation des banques différentielles d'épissage de l'invention sont illustrées notamment sur les Figures 13 à 15. Ainsi, ces banques sont utilisables pour :

##### 4.1. L'évaluation du potentiel toxique d'un composé (figure 13) :

Dans cet exemple, la situation de référence est désignée A et la situation toxique est désignée B. Des abaques de toxicité sont obtenues par

traitement de la situation A en présence de différentes concentrations d'un composé toxique de référence, pendant des périodes variables. A différents points les abaques de toxicité, des banques différentielles qualitatives sont construites (paires de banque), dans cet exemple, des banques restreintes rA/cB et rB/cA. Les paires de banque sont avantageusement déposées sur un support. Le support est ensuite hybridé avec des sondes issues de l'échantillon biologique initial traité par différentes doses de composés test : Produits X, Y et Z. L'hybridation est révélée et fait apparaître le potentiel toxique des produits test : dans cet exemple, le produit Z présente une forte toxicité et le produit Y offre un profil intermédiaire. La faisabilité de cette constitution d'abaques de toxicité est bien illustrée par l'exemple de constitution de banques de criblage différentiel qualitatif décrit ci-avant et mettant en jeu l'éthanol et des cellules HepG2.

4.2. L'évaluation de l'efficacité d'une composition pharmaceutique (figure 14) :

Dans cet exemple, une paire de banques restreintes selon l'invention est réalisée à partir d'un modèle pathologique B et d'un modèle sain A (ou du modèle pathologique traité avec un produit actif de référence). Les banques différentielles rA/cB et rB/cA sont, le cas échéant, déposées sur un support. Cette paire de banque regroupe les différences d'épissage entre les deux situations. Cette paire de banque permet d'évaluer l'efficacité d'un composé test, c'est-à-dire de déterminer sa capacité à générer un profil de type "sain" (rA/cB) à partir du profil de type pathologique (rB/cA). Dans cet exemple, la paire de banque est hybridée avec des sondes préparées à partir des situations A et B avec ou sans traitement par le composé test. Le profil d'hybridation qui peut être obtenu est présenté sur la figure 14. La faisabilité de cette application est la même que celle de la constitution de banques de différences qualitatives caractéristiques de situations saines et toxiques présentée plus-haut. La situation toxique est remplacée par l'état pathologique et il est possible d'apprécier la capacité d'un composé test à

produire une sonde hybridant avec plus ou moins de préférence avec les situations de référence ou pathologique.

4.3. L'anticipation de la réponse d'un échantillon pathologique à un  
5 traitement (figure 15) :

Dans cet exemple, une paire de banque restreinte selon l'invention est réalisée à partir de deux modèles pathologiques, dont l'un répond à un traitement par un produit donné (le gène p53 sauvage par exemple) : situation A ; et l'autre y est réfractaire : situation B. Cette paire de  
10 banque (rA/cB ; rB/cA) est déposée sur un support.

Cette paire de banque est ensuite utilisée pour déterminer la sensibilité d'un échantillon pathologique test à ce même produit. Pour cela, cette paire de banque est hybridée avec des sondes provenant de biopsies de patients dont on souhaite anticiper la réponse au traitement de référence.  
15 Le profil d'hybridation d'une biopsie de répondeur et d'une biopsie de non-répondeur est présenté sur la figure 15.

#### 4.4 L'identification de ligands pour des récepteurs orphelins

L'activation de récepteurs membranaires ou nucléaires par  
20 leurs ligands pourrait induire spécifiquement des dysrégulations dans l'épissage de certains ARNs. L'identification de ces événements par les méthodes DATAS de l'invention permet de disposer d'un outil (marqueurs, banques, kits, etc.) de suivi d'activation de récepteurs, utilisables pour la recherche de ligands naturels ou synthétiques de récepteurs, en particulier  
25 orphelins. Selon cette application, des marqueurs associés aux dysrégulations sont identifiés et déposés sur des supports. L'ARN total de cellules, (sur)exprimant le récepteur à l'étude, traitées ou non par différentes compositions et/ou composés tests, est extrait et utilisé comme sonde dans une hybridation avec les supports. La mise en évidence d'une hybridation  
30 avec certains, voire la totalité des marqueurs déposés sur le support, indique que le récepteur à l'étude a été activé, et donc que le composition/composé

correspondant constitue ou comprend un ligand dudit récepteur.

#### 4.5 L'identification de cibles d'intérêt thérapeutique :

Celle-ci se fait par identification de gènes dont l'épissage est  
5 modifié dans une pathologie ou dans un modèle de pathologie et plus  
précisément l'identification des exons ou introns modifiés. Cette approche  
doit permettre de donner accès aux séquences qui codent les domaines  
fonctionnels altérés lors de pathologies ou de tout phénomène  
physiopathologique adressant les phénomènes de prolifération,  
10 différenciation ou apoptose par exemple.

Un exemple de l'apport du criblage différentiel qualitatif à  
l'identification de gènes différentiellement épissés est fourni par l'application  
de DATAS à un modèle d'induction d'apoptose par induction de l'expression  
de la forme sauvage de p53. Ce modèle cellulaire a été établi par  
15 transfection d'un système d'expression inductible pour le gène suppresseur  
de tumeur p53. De façon à identifier les différences qualitatives qui sont  
associés spécifiquement à l'apoptose induite par p53, DATAS a été mise en  
œuvre à partir des ARN messagers extraits de cellules induites et non  
induites. Pour ces expériences 200ng d'ARN polyA+ et 200ng d'ADNc ont  
20 été utilisés lors de la formation des hétéroduplex. Une centaine de clones a  
été obtenue à partir de chacune des hybridations croisées. L'hybridation de  
ces clones bactériens puis des fragments d'ADNc qu'ils contiennent avec  
des sondes représentatives des ARN messagers totaux des situations de  
départ a permis l'identification de séquences spécifiquement exprimées lors  
25 de la forte induction de p53 qui aboutit à la mort cellulaire (figure 16).

Ces fragments dérivent de séquences exoniques ou introniques qui  
modulent la qualité du message présent et permettent de proposer les  
domaines fonctionnels auxquels ils participent ou qu'ils interrompent comme  
des cibles d'intervention pour induire ou inhiber la mort cellulaire.

30 Une telle approche aboutit également à la constitution d'une paire de  
banques qui rassemblent des événements différentiellement épissés entre

une situation non apoptotique et une situation apoptotique. Cette paire de banque peut être utilisée pour tester le pouvoir hybridant d'une sonde dérivée d'une autre situation physiopathologique ou d'un traitement particulier. Le résultat d'une telle hybridation donnera des indications sur l'engagement éventuel du programme d'expression génétique de la situation testée vers l'apoptose.

Comme il ressort de la description ci-avant, l'invention concerne également :

- 10 - toute sonde nucléique, tout oligonucléotide, tout anticorps dirigé contre une séquence identifiée par la technique décrite dans la présente demande et caractérisés en ce qu'ils permettent de caractériser une situation pathologique,
- l'utilisation des informations issues de l'utilisation des techniques décrites pour la recherche de molécules organiques à visée thérapeutique par la mise en place de criblages caractérisés en ce qu'ils ciblent des domaines splicés différenciellement entre une situation saine et une situation pathologique ou bien caractérisés en ce qu'ils sont basés sur l'inhibition des gains de fonctions acquis par la protéine résultant d'un épissage différentiel,
- l'utilisation des informations issues des techniques décrites dans la présente demande pour des applications de thérapie génique,
- l'utilisation d'ADNc transférés par thérapie génique caractérisés en ce qu'ils ont des propriétés antagonistes ou agonistes sur des voies de signalisation cellulaires définies,
- toute constitution et toute utilisation de banques moléculaires d'exons ou d'introns alternatifs à des fins:
  - . de diagnostic ou de réactifs commerciaux pour la recherche
  - 30 . de créer ou de rechercher des molécules, polypeptides, acides nucléiques pour application thérapeutique.

- toute constitution et toute utilisation de banques virtuelles informatiques regroupant des exons ou introns alternatifs caractérisés en ce que ces banques permettent de concevoir des sondes nucléiques ou des amorces oligonucléotidiques dans le but de caractériser les épissages alternatifs qui différentient deux états physiopathologiques distincts.

- toute composition pharmaceutique ou diagnostique comprenant des polypeptides, acides nucléiques sens ou anti-sens, ou des molécules chimiques capables d'interférer avec les produits d'épissage alternatifs mis en évidence et clonés par les techniques de l'invention,

toute composition pharmaceutique ou diagnostique comprenant des polypeptides, acides nucléiques sens ou anti-sens, ou des molécules chimiques capables de restaurer un épissage représentatifs d'une situation normale par opposition à l'événement alternatif caractéristique d'une situation pathologique.

15

## **5. DEREGULATIONS DES MÉCANISMES D'ÉPISSAGE DES ARNS PAR DES AGENTS TOXIQUES.**

Cet exemple montre que les différences de formes et/ou profils d'épissages peut être utilisée comme marqueur pour le suivi et/ou la détection de toxicité et/ou d'efficacité de composés.

Les effets d'agents toxiques sur les dérégulations d'épissages des ARNs a été testé de la manière suivante. Des cellules hépatocytaires, HepG2, ont été traitées par différentes doses de trois composés toxiques (ethanol, camptothecine, PMA (Phorbol 12-Myristate 13-Acetate)). Deux tests de cytotoxicité (Bleu de Trypan, MTT) ont été réalisés à différents temps : 4h et 18h pour l'ethanol ; 4h et 18h pour la camptothecine ; 18h et 40h pour le PMA.

Le Bleu de Trypan est un colorant qui peut être incorporé par les cellules vivantes. Un simple comptage des cellules "bleues" et "blanches" sous microscope permet de déterminer le pourcentage de cellules vivantes

après traitement ou pourcentage de survie. Les points sont effectués en triplicates.

Le test MTT est un test colorimétrique qui mesure la capacité des cellules vivantes à convertir les sels solubles de tetrazolium (MTT) en un précipité insoluble de formazan. Ces cristaux de formazan, bleu foncé, peuvent être dissous et leur concentration déterminée par mesure d'absorbance à 550 nm. Ainsi, après ensemencement de plaques 24 puits par 150000 cellules sur la nuit, puis traitement des cellules par les composés toxiques, est ajouté 50 µl de MTT (Sigma) (à 5 mg/ml dans du PBS). La réaction de formation des cristaux de formazan s'effectue en 5 h dans l'incubateur à CO<sub>2</sub> (37°C, 5% CO<sub>2</sub>, 95% humidité). Après addition de 500 µl de solution de solubilisation (Hcl 0,1N dans Isopropanol-Triton X-100 (10%)), les cristaux sont dissous sous agitation et les absorbances mesurées à 550 à 660nm. Les points sont effectués en triplicates avec les contrôles (viabilité, mort cellulaire, blancs) appropriés.

Un test d'apoptose ou mort cellulaire programmée a également été réalisé par mesure de la fragmentation d'ADN via l'utilisation d'anticorps anti-histone et de mesures d'ELISA. Le test utilisé est Cell Death ELISA Plus de Roche.

Les résultats de ces trois tests (Figures 18 A,B,C) ont permis de déterminer que les doses suivantes :

- ethanol : 0,1 M
  - camptothecine : 1 µg/ml
  - PMA : 50 ng/ml
- étaient bien inférieures aux IC<sub>50</sub>s mesurées.

Les cellules HepG2 ont ainsi été traitées par ces trois composés à ces trois doses pendant 4 h pour l'ethanol et la camptothecine et 18 h pour le PMA. Les ARNs messagers ont été purifiés par billes Dynal-Oligo-(dT) à partir d'ARNs totaux purifiés selon le kit Rneasy (Quiagen). Des synthèses d'ADNc ont été effectuées à partir de ces ARNs messagers et de la

Transcriptase Inverse Superscript (Life Technologies) en utilisant des hexamères aléatoires comme amorces.

Ces premiers brins ont servi de matrices à des réactions d'amplification par PCR (94°C 1mn, 55°C 1mn, 72°C 1mn, 30 cycles) à l'aide  
5 des amorces oligonucléotidiques suivantes :

MACH- $\alpha$ :

5'-TGCCCAAATCAACAAGAGC-3' (SEQ ID NO: 11)

10 5'-CCCCTGACAAGCCTGAATA-3' (SEQ ID NO: 12)

Ces amorces correspondent à des régions communes aux différentes isoformes décrites de MACH- $\alpha$  (1,2 et 3, amplifiant respectivement 595, 550 et 343 paires de bases). MACH- $\alpha$  (Caspase-8) est une protéase impliquée  
15 dans la mort cellulaire programmée (Boldin et.al., Cell (1996), 85, 803-815).

BCL-X :

5' ATGTCTCAGAGCAACCGGGAGCTG 3' (SEQ ID NO: 13)

20 5' GTGGCTCCATTCACCGCGGGGCTG 3' (SEQ ID NO: 14)

Ces amorces correspondent à des régions communes aux différentes isoformes décrites de bcl-X (bcl-XI, bcl-Xs, BCL-X $\beta$ ) (Boise et al., Cell (1993) 74, 597-608; U72398 (Genbank)) et doivent amplifier un fragment unique  
25 pour ces trois isoformes de 204 paires de bases.

FASR:

5'-TGCCAAGAAGGGAAGGAGT-3' (SEQ ID NO: 15)

30 5'-TGTCATGACTCCAGCAATAG-3' (SEQ ID NO: 16)



Ces amorces correspondent à des régions communes à certaines isoformes de FASR et doivent amplifier un fragment de 478 paires de bases pour la forme sauvage de FasR, 452 pour l'isoforme  $\Delta 8$  et 415 pour l'isoforme  $\Delta TM$ .

Les résultats rapportés sur la figure 19 indiquent que :

- la camptothécine induit une diminution de l'expression de l'isoforme MACH- $\alpha 1$  et une augmentation de l'isoforme MACH- $\alpha 3$ .
- 10 • La camptothécine induit l'apparition d'une nouvelle isoforme de bcl-X (bande supérieure du doublet migrant vers 200 paires de bases).
- La camptothécine induit une diminution de la forme sauvage du récepteur de fas, remplacé par une expression d'une isoforme plus courte pouvant correspondre à Fas  $\Delta TM$ .
- 15 • L'éthanol induit la disparition de bcl-x, remplacé par une isoforme plus courte.
- L'éthanol induit une augmentation de la forme longue, sauvage, du récepteur de fas, ceci aux dépens de l'isoforme plus courte.

20 Ces résultats démontrent que des traitements par des agents toxiques à des doses faibles peuvent induire des dysrégulations d'épissages alternatifs de certains ARNs, ceci de façon spécifique. L'identification de ces dysrégulations au niveau post-transcriptionnel, notamment par l'application de la technologie DATAS permet ainsi de définir un outil prédictif de la

25 toxicité de molécules.

## REVENDEICATIONS

1. Procédé d'identification et/ou de clonage de régions d'acides nucléiques représentatives de différences génétiques qualitatives entre deux échantillons biologiques, comprenant une étape d'hybridation d'une population d'ARN ou d'ADNc double-brin provenant d'un premier échantillon biologique avec une population d'ADNc provenant d'un deuxième échantillon biologique.

2. Procédé selon la revendication 1 caractérisé en ce qu'il comprend :

(a) l'hybridation des ARN provenant du premier échantillon (situation test) avec les ADNc provenant du deuxième échantillon (situation de référence);

(b) l'hybridation des ARN provenant du deuxième échantillon (situation de référence) avec les ADNc provenant du premier échantillon (situation test); et

(c) l'identification et/ou le clonage, à partir des hybrides formés en (a) et (b), d'acides nucléiques correspondant à des différences génétiques qualitatives.

3. Procédé selon la revendication 1 ou 2 caractérisé en ce que les hybridations sont réalisées entre une population d'ARN et des ADNc simple-brins et en ce qu'il comprend l'identification et/ou le clonage de régions d'ARN non appariées.

4. Procédé selon la revendication 1 ou 2 caractérisé en ce que les hybridations sont réalisées entre une population d'ARN et des ADNc double-brins et en ce qu'il comprend l'identification et/ou le clonage de régions d'ADN appariées.

5. Procédé selon la revendication 1 caractérisé en ce qu'il comprend

l'hybridation entre une population d'ADNc double-brin provenant d'un premier échantillon biologique et une population d'ADNc simple-brin provenant d'un deuxième échantillon biologique.

5           6. Procédé selon l'une des revendications 1 à 5 caractérisé en ce que l'échantillon biologique est composé de cellules, d'un tissu, d'un organe ou d'une biopsie.

7. Procédé selon l'une des revendications 1 à 6 pour l'identification  
10 et/ou le clonage d'épissages alternatifs différentiels entre des cellules tumorales et des cellules non-tumorales.

8. Procédé selon l'une des revendications 1 à 6 pour l'identification  
15 et/ou le clonage d'épissages alternatifs différentiels entre des cellules traitées par un composé test et des cellules non-traitées.

9. Procédé selon l'une des revendications 1 à 6 pour l'identification  
et/ou le clonage d'épissages alternatifs différentiels entre des cellules en  
apoptose et des cellules non-apoptotiques.

20

10. Procédé selon l'une des revendications 1 à 9 caractérisé en ce que l'hybridation est réalisée en phase liquide.

11. Procédé d'identification et/ou de clonage de régions d'acides  
25 nucléiques épissées différenciellement entre deux situations physiologiques A et B, comprenant :

(a) la formation d'hétéroduplex en phase liquide entre les ARN messagers provenant de la situation A et les ADNc provenant de la situation B d'une part;

30           (b) la formation d'hétéroduplex en phase liquide entre les ARN messagers provenant de la situation B et les ADNc provenant de la situation

A d'autre part; et

(c) l'identification et/ou le clonage des régions d'ARN non appariées dans les hétéroduplex obtenus en (a) et en (b).

5 12. Procédé d'identification et/ou de clonage de régions d'acides nucléiques épissées différemment entre deux situations physiologiques A et B, comprenant :

(a) la formation d'hétéroduplex entre les ARN messagers provenant de la situation A et les ADNc provenant de la situation B d'une  
10 part, les ARN ou les ADNc étant immobilisés sur un support ;

(b) la formation d'hétéroduplex entre les ARN messagers provenant de la situation B et les ADNc provenant de la situation A d'autre part, les ARN ou les ADNc étant immobilisés sur un support ; et

(c) l'identification et/ou le clonage des régions d'ARN non  
15 appariées dans les hétéroduplex obtenus en (a) et en (b).

13. Composition comprenant les acides nucléiques identifiés et/ou clonés selon les procédés des revendications 1 à 12.

20 14. Composition d'acides nucléiques, caractérisée en ce qu'elle comprend essentiellement les acides nucléiques représentatifs d'altérations génétiques qualitatives, notamment des épissages alternatifs distinguant deux situations physiologiques d'une cellule ou d'un tissu.

25 15. Composition selon les revendications 13 ou 14 caractérisée en ce que les acides nucléiques sont clonés dans des vecteurs.

30 16. Banque d'acides nucléiques comprenant des acides nucléiques spécifiques d'altérations génétiques qualitatives, notamment d'épissages alternatifs distinguant deux situations physiologiques d'une cellule ou d'un tissu.

17. Banque selon la revendication 16, caractérisée en ce qu'il s'agit d'une banque restreinte aux épissages alternatifs qui caractérisent les ARN matures.

5

18. Banque selon la revendication 16, caractérisée en ce qu'il s'agit d'une banque complexe des épissages alternatifs qui caractérisent les transcrits.

10

19. Banque selon la revendication 16, caractérisée en ce qu'il s'agit d'une banque autologue caractéristique des épissages alternatifs entre les ARN matures et pré-messagers d'une situation physiologique.

20. Banque d'acides nucléiques comprenant des oligonucléotides ou des fragments PCR spécifiques d'épissages alternatifs distinguant deux situations physiologiques.

21. Banque de microorganismes comprenant des microorganismes transformés par des acides nucléiques spécifiques d'épissages alternatifs distinguant deux situations physiologiques d'une cellule ou d'un tissu.

22. Banque selon les revendications 16 à 21 caractérisée en ce qu'elle est déposée sur un support.

23. Kit comprenant un support sur lequel est déposé une banque selon l'une des revendications 16 à 21.

24. Kit selon la revendication 23 caractérisé en ce qu'il comprend deux banques selon l'une des revendications 16 à 21, déposée sur un même support ou sur deux supports distincts.

30

25. Kit selon les revendications 23 ou 24 caractérisé en ce que le support est composé d'un filtre, membrane ou d'une puce.

26. Utilisation d'une composition selon l'une des revendications 13 à 15 ou d'une banque selon les revendications 16 à 21 pour l'identification de molécules actives.

27. Utilisation d'une composition selon l'une des revendications 13 à 15 ou d'une banque selon les revendications 16 à 21 pour l'identification de protéines ou domaines protéiques affectés dans une pathologie.

28. Utilisation d'une composition selon l'une des revendications 13 à 15 ou d'une banque selon les revendications 16 à 21 pour l'identification de domaines antigéniques spécifiques de protéines impliquées dans une pathologie.

29. Méthode d'identification et/ou de production de protéines ou domaines protéiques impliqués dans une pathologie comprenant :

(a) l'hybridation des ARN messagers d'un échantillon pathologique avec les ADNc d'un échantillon sain, ou l'inverse, ou les deux en parallèle,

(b) l'identification, dans des hybrides formés, des régions correspondant aux différences qualitatives spécifiques de l'état pathologique par rapport à l'état sain,

(c) l'identification et/ou la production d'une protéine ou domaine protéique correspondant à une région identifiée en (b).

30. Procédé d'identification et/ou de clonage de gènes suppresseurs de tumeurs ou d'épissages au sein de gènes suppresseurs de tumeurs, comprenant :

(a) l'hybridation des ARN messagers d'un échantillon de tumeur

avec les ADNc d'un échantillon sain, ou l'inverse, ou les deux en parallèle,

(b) l'identification, dans les hybrides formés, des régions correspondant aux différences qualitatives spécifiques de l'échantillon tumoral par rapport à l'état sain,

5 (c) l'identification et/ou la production d'une protéine ou domaine protéique correspondant à une région identifiée en (b).

31. Composition comprenant un composé capable d'interférer avec les produits d'épissages alternatifs identifiés selon le procédé des revendications  
10 1 à 12.

32. Protéine susceptible d'être identifiée par le procédé de la revendication 29.

15 33. Utilisation d'une banque selon les revendications 16 à 21 ou d'un kit selon les revendications 23 à 25 pour évaluer la toxicité d'un composé.

34. Procédé d'identification et/ou de clonage d'acides nucléiques spécifique d'un état toxique d'un échantillon biologique donné comprenant la  
20 préparation de banques différentielles qualitatives entre les ARN et les ADNc de l'échantillon après ou sans traitement par un composé toxique test, et la recherche de marqueurs de toxicité spécifiques des qualités de l'échantillon après traitement.

25 35. Procédé de détermination ou d'évaluation de la toxicité d'un composé test sur un échantillon biologique donné comprenant l'hybridation :

- de banques différentielles entre les ADNc et les ARN dudit échantillon biologique à l'état sain et à un ou différents stades de toxicité résultant d'un traitement dudit échantillon avec un composé toxique de  
30 référence, avec,

- une préparation d'acides nucléiques de l'échantillon biologique

traité par ledit composé test, et

- l'évaluation du potentiel toxique du composé test par analyse du degré d'hybridation avec les différentes banques.

5           36. Procédé selon la revendication 35 caractérisé en ce que l'échantillon biologique est une culture d'hépatocytes, de cellules épithéliales rénales ou de cellules endothéliales, traitée ou non par un agent toxique, de préférence l'éthanol.

10           37. Procédé selon la revendication 35 caractérisé en ce que l'échantillon biologique est une culture de peau traitée ou non par des agents toxiques ou irritants.

            38. Utilisation d'une banque selon les revendications 16 à 21 ou d'un  
15   kit selon les revendications 23 à 25 pour évaluer l'efficacité d'un composé.

            39. Procédé de détermination ou d'évaluation de l'efficacité thérapeutique d'un composé test sur un échantillon biologique donné comprenant l'hybridation :

20           - de banques différentielles entre les ADNc et les ARN dudit échantillon biologique à l'état sain et à l'état pathologique avec,  
            - une préparation d'acides nucléiques de l'échantillon biologique traité par ledit composé test, et  
            - l'évaluation du potentiel thérapeutique du composé test par  
25   analyse du degré d'hybridation avec les différentes banques.

            40. Utilisation d'une banque selon les revendications 16 à 21 ou d'un kit selon les revendications 23 à 25 pour évaluer la réponse d'un échantillon pathologique à un composé.

30

            41. Procédé de détermination ou d'évaluation de la réponse d'un



patient à un composé ou traitement test comprenant l'hybridation :

- de banques différentielles entre les ADNc et les ARN d'un échantillon biologique répondeur audit composé/traitement et d'un échantillon biologique non-répondeur ou mal-répondeur audit composé/traitement, avec,
- une préparation d'acides nucléiques d'un échantillon biologique pathologique du patient, et
- l'évaluation du potentiel répondeur du patient par analyse du degré d'hybridation avec les différentes banques.

10

42. Procédé selon la revendication 41 pour la détermination ou l'évaluation de la réponse d'un patient à un composé ou traitement antitumoral.

15

43. Procédé selon la revendication 42 pour la détermination ou l'évaluation de la réponse d'un patient à un traitement antitumoral par transfert du gène p53 sauvage.

44. Acide nucléique susceptible d'être identifié par le procédé selon les revendications 1 à 12.

20

45. Utilisation d'un acide nucléique selon la revendication 44 pour la détection d'anomalies génétiques dans un échantillon.

25

46. Utilisation d'un composé selon la revendication 28 pour la détection d'une anomalie génétique dans un échantillon.

47. Anticorps dirigé contre une protéine ou un domaine protéique tel que défini dans la revendication 27 ou 28.

30

48. Protéine  $\Delta$ SHC de séquence SEQ ID NO: 9.

49. Sonde nucléique, oligonucléotide ou anticorps permettant d'identifier la protéine  $\Delta$ SHC selon la revendication 48 ou son acide nucléique, et/ou une altération du rapport SHC/ $\Delta$ SHC dans un échantillon biologique.

5

50. Procédé de criblage caractérisé en ce qu'il est basé sur le blocage du domaine épissé dans la protéine SHC ou sur le blocage des gains de fonction acquis par la protéine épissée  $\Delta$  SHC.

10 51. Vecteur comprenant une séquence codant pour la protéine  $\Delta$ SHC selon la revendication 48.

52. Procédé selon les revendications 3, 4 ou 5 caractérisé en ce que le clonage des acides nucléiques comprend la transcription inverse et/ou  
15 l'amplification au moyen d'amorces aléatoires ou semi-aléatoires, en particulier d'amorces de séquence SEQ ID NO: 3 dans laquelle N indique que chacune des quatre bases peut être présente de façon aléatoire à la position indiquée, W, X et Y désignent chacun une base déterminée, et Z désigne soit une base déterminée, soit un groupe 3'-OH.

20

53. Oligonucléotide comprenant, dans l'orientation 5'  $\rightarrow$  3' :

- une zone stabilisatrice comprenant 8 à 24 nucléotides déterminés,
- une région aléatoire comprenant de 3 à 8 nucléotides, et
- une zone minimale d'amorçage comprenant 2 à 4 nucléotides

25 définis.

54. Oligonucléotide de séquence SEQ ID NO: 3 dans laquelle :

- N indique que chacune des quatre bases peut être présente de façon aléatoire à la position indiquée;

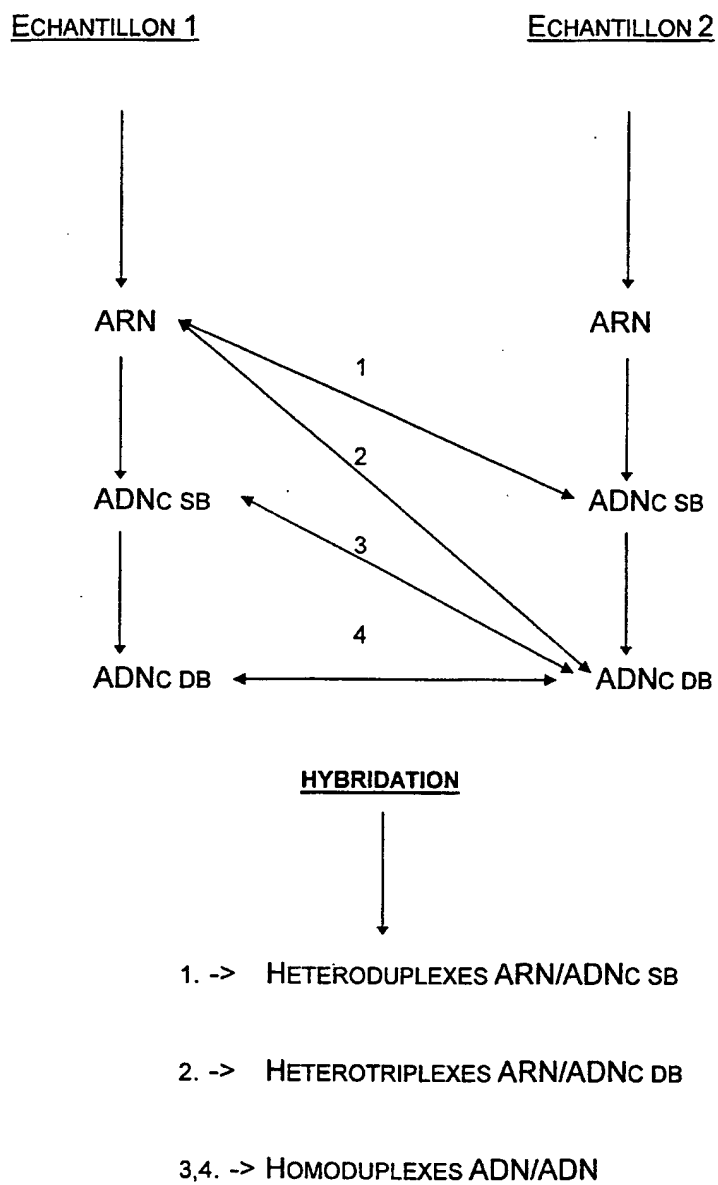
- W, X et Y désignent chacun une base déterminée,

30

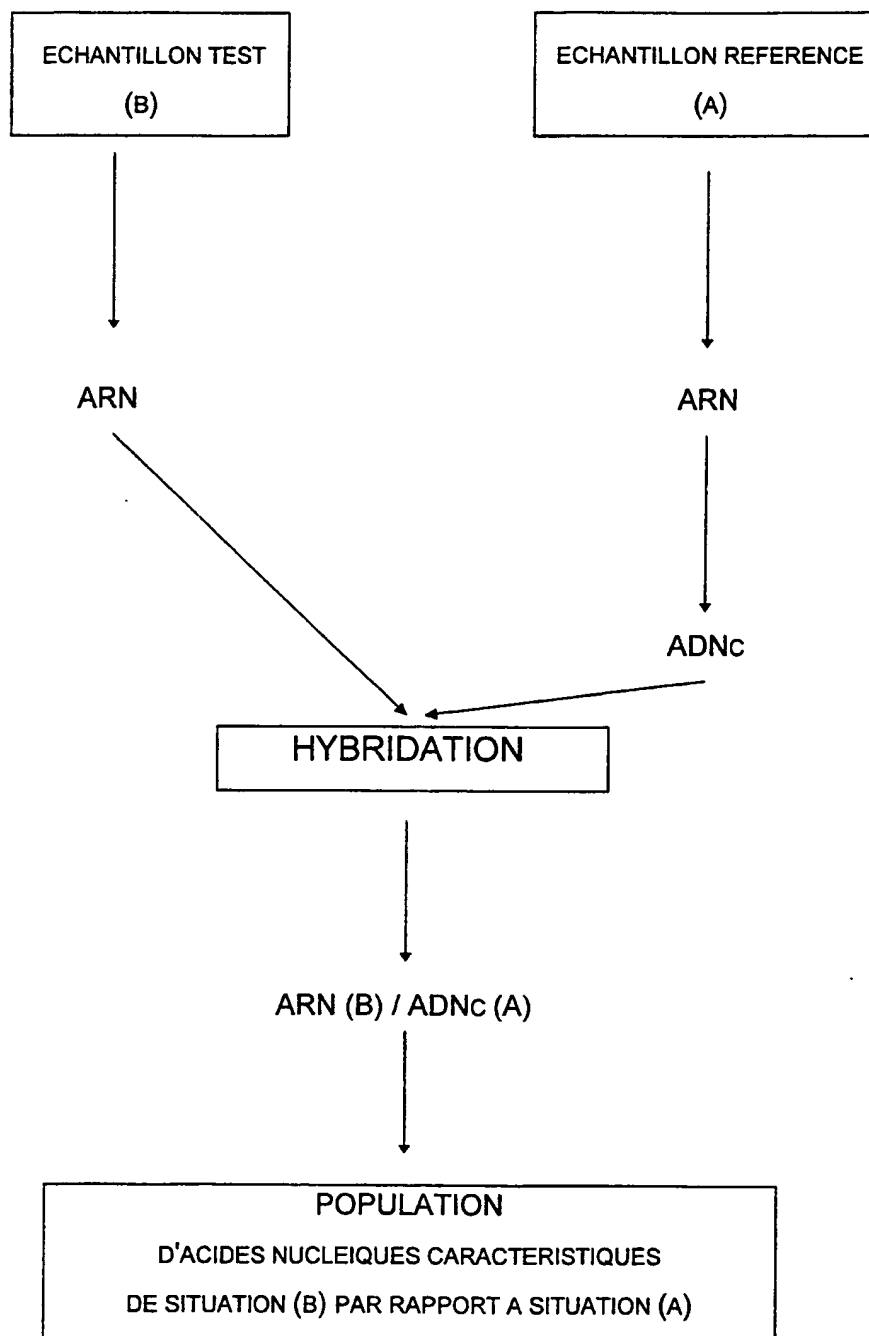
- Z désigne soit une base déterminée, soit un groupe 3'-OH.

55. Banque d'ADN génomique caractérisée en ce qu'elle est constituée de fragments d'ADN génomiques de taille inférieure ou égale à 1 kb environ.
- 5
56. Méthode de détection ou de suivi du potentiel toxique et/ou thérapeutique d'un composé, basée sur la détection de formes et/ou de profils d'épissages induits par ce composé sur un échantillon biologique.
- 10
57. Utilisation, comme source de marqueurs de pharmacogénomique, (i) de l'intervariabilité, entre individus, d'isoformes générées par épissage alternatif (analyse du spliceome) ou (ii) de modifications d'épissage induites par des traitements.

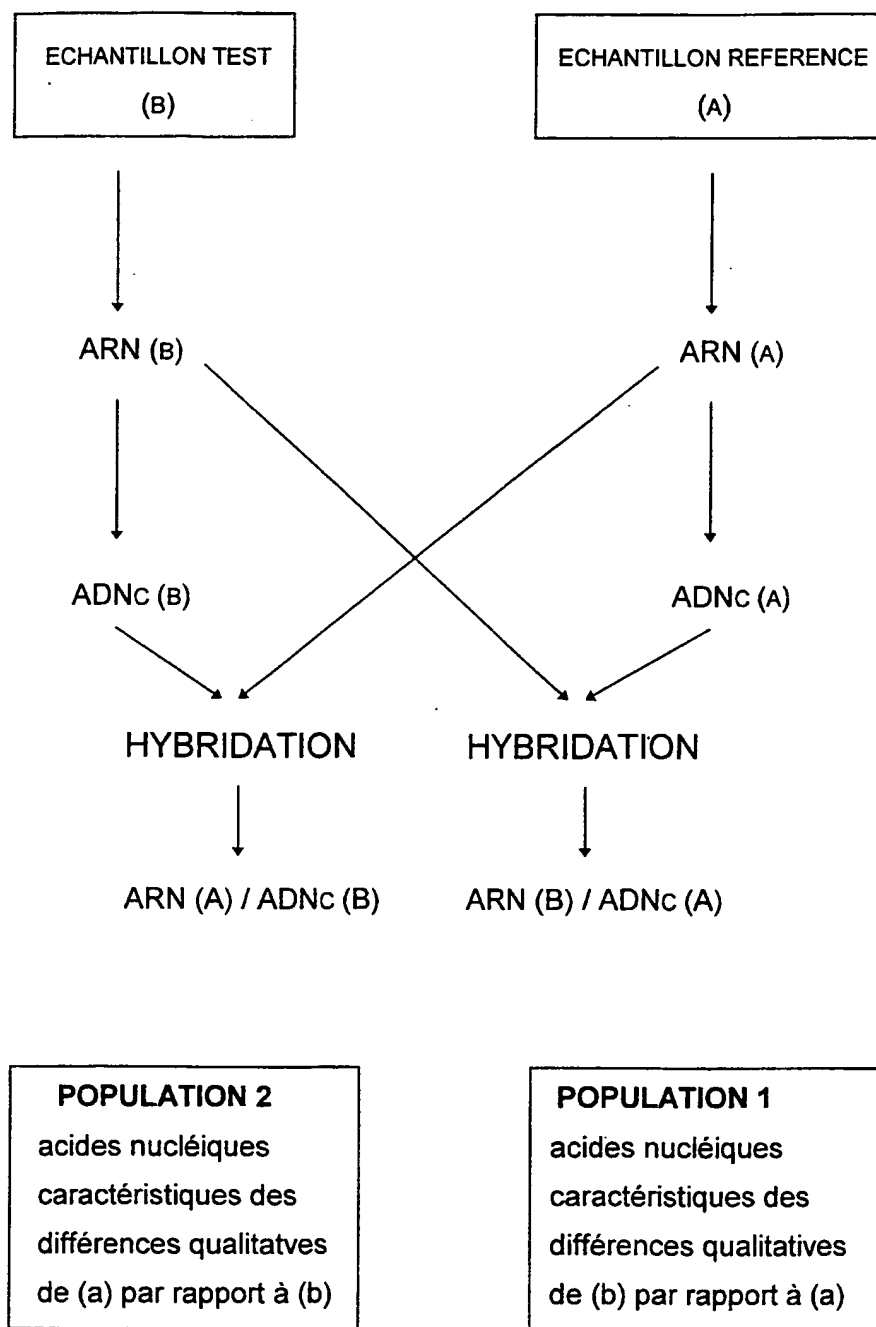
1/26

FIGURE 1A

2/26

**FIGURE 1B**

3/26

**FIGURE 1C**

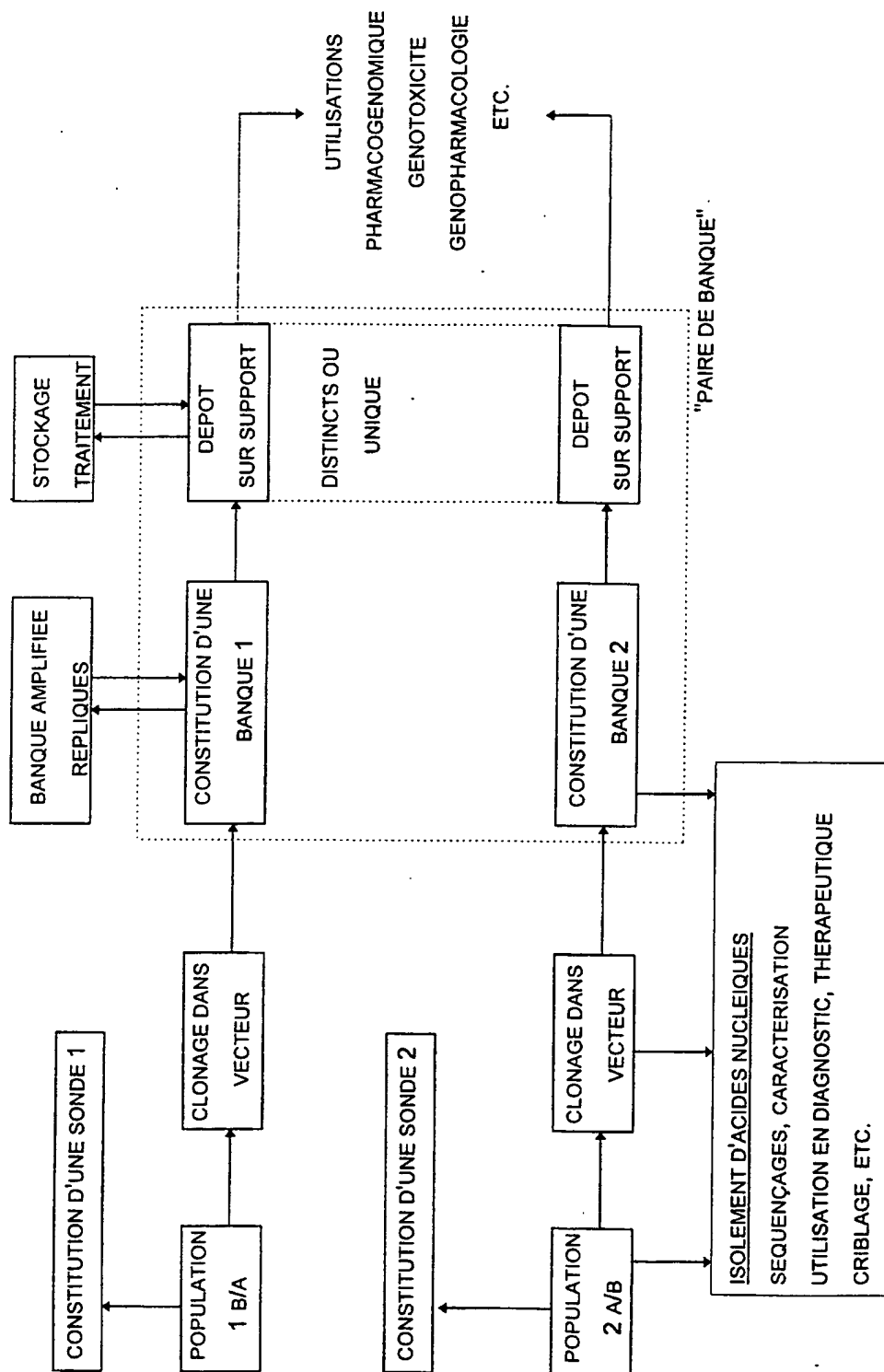
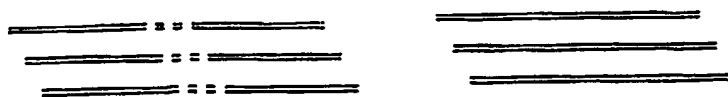


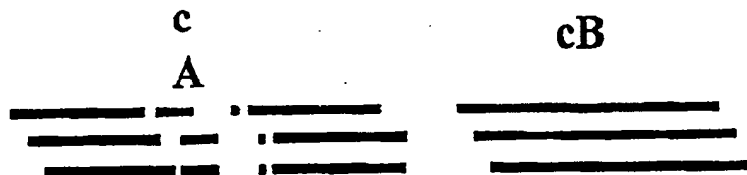
FIGURE 1D

5/26

ARN (mA)                      ARN (mB)  
Echantillon pathologique      Echantillon normal



ADN complémentaire



Hybrides ARN<sub>patho</sub> / cDNA<sub>normal</sub> (mA/cB)



**FIGURE 2**



6/26

Hybrides  $ARN_{patho}$  /  $cDNA_{normal}$



Séquence non épissée après digestion à la RNaseH

=====



Séquence recherchée marquée en 5' et 3'  
à l'aide de deux oligonucléotides

■ ■ ■ = = = ■ ■ ■



Fragment amplifié par PCR

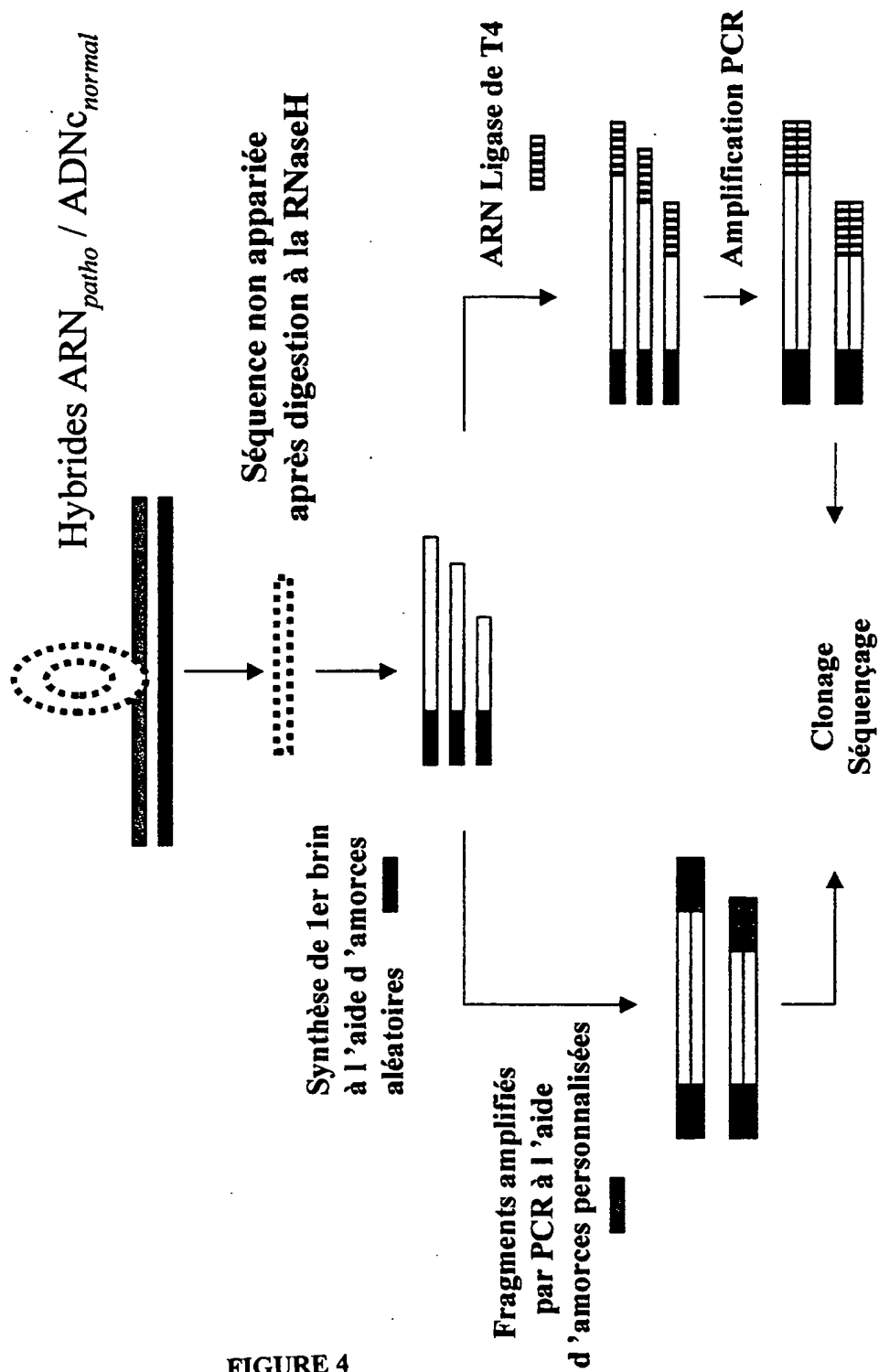
■ ■ ■ — ■ ■ ■



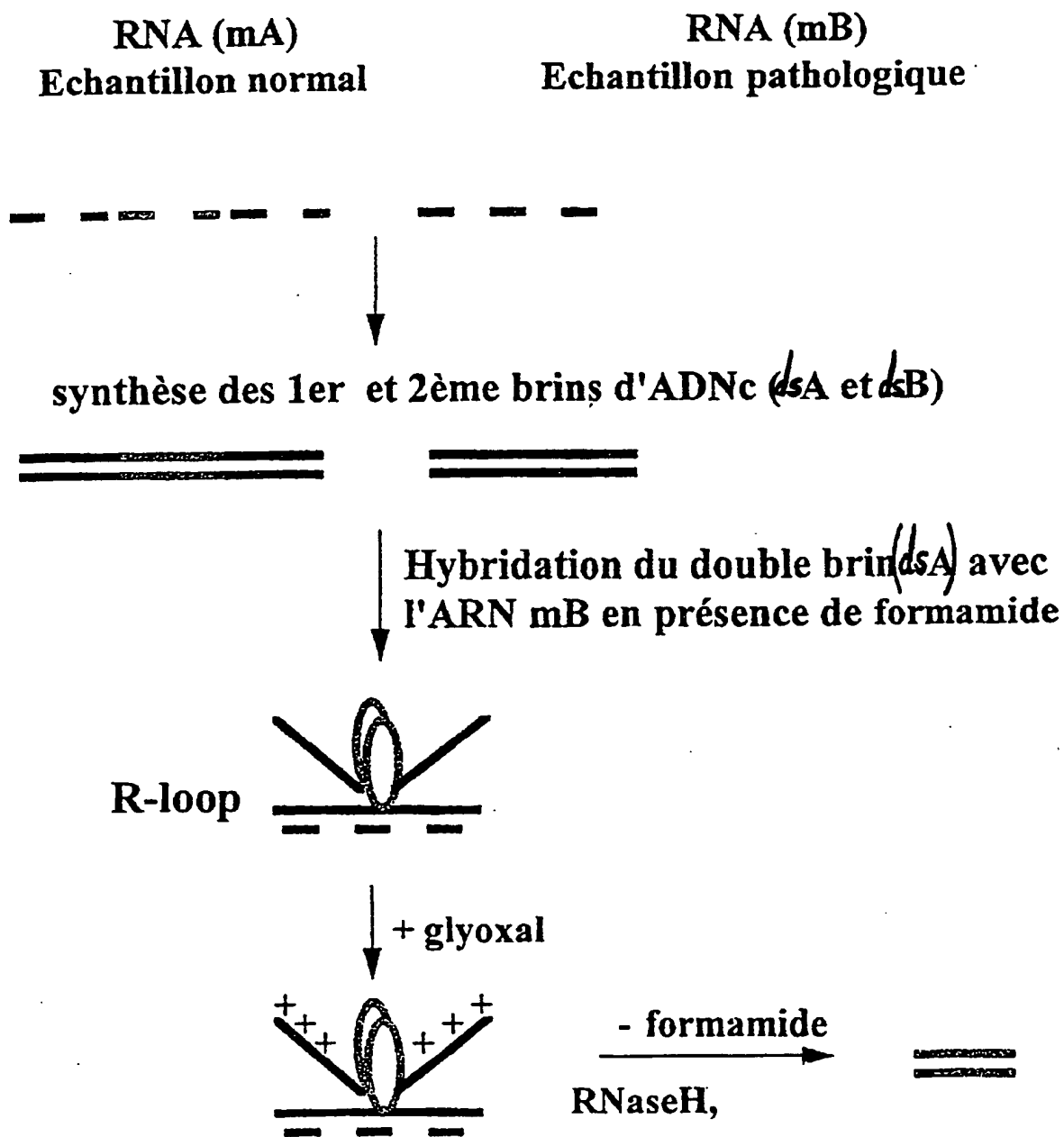
Clonage et séquençage

**FIGURE 3**

7/26



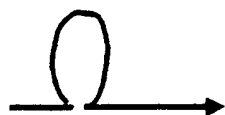
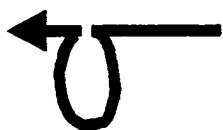
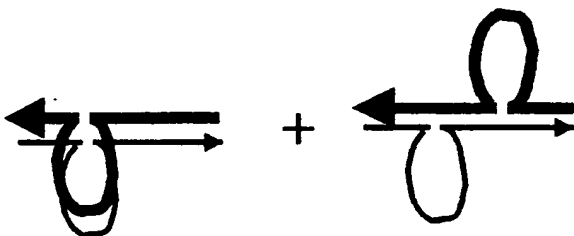
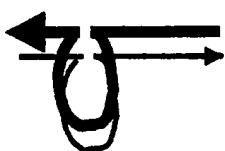
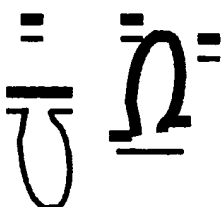
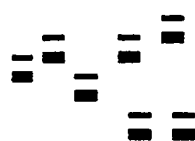
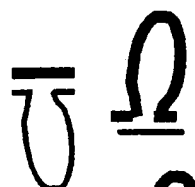
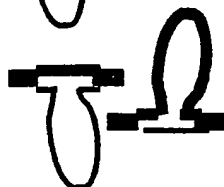
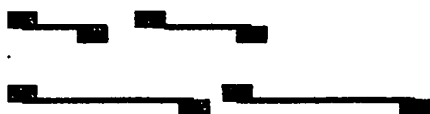
**FIGURE 4**



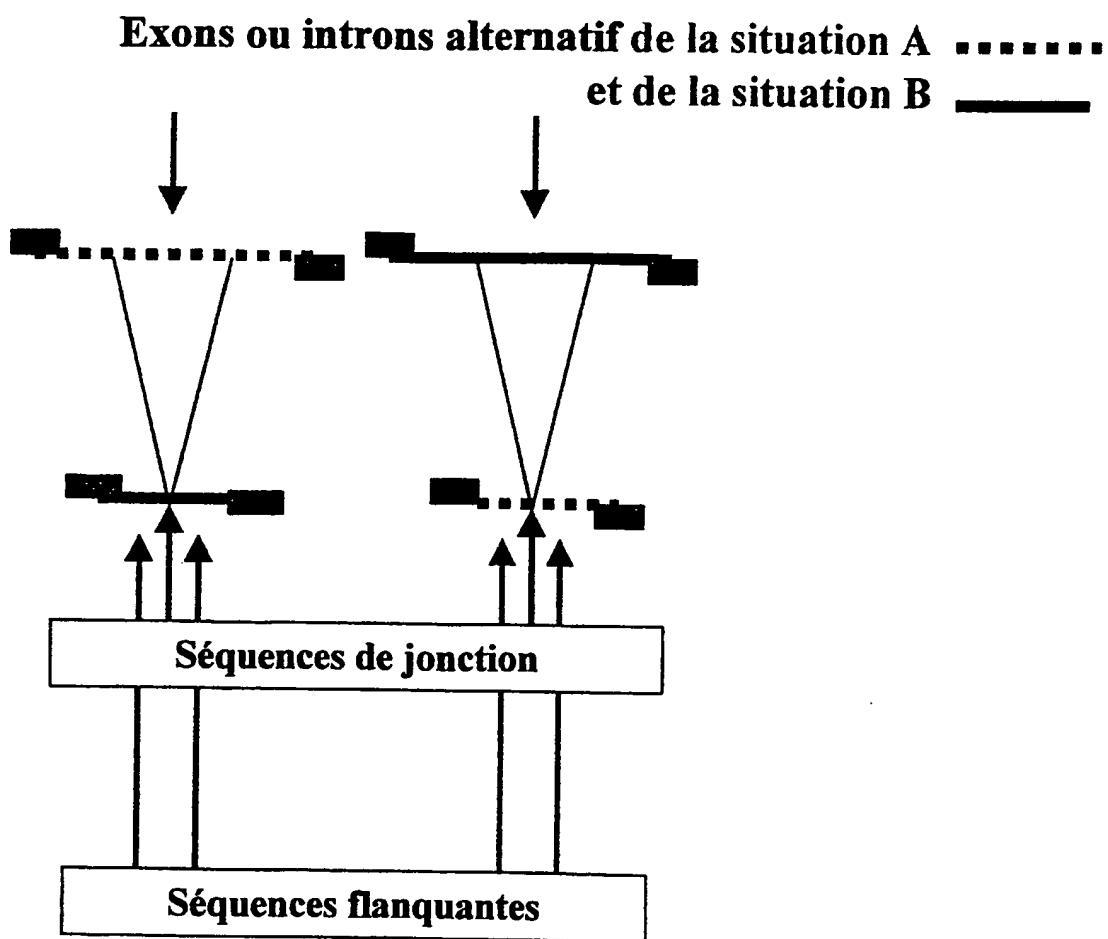
**FIGURE 5**

**MUNG BEAN NUCLEASE  
S 1 NUCLEASE**

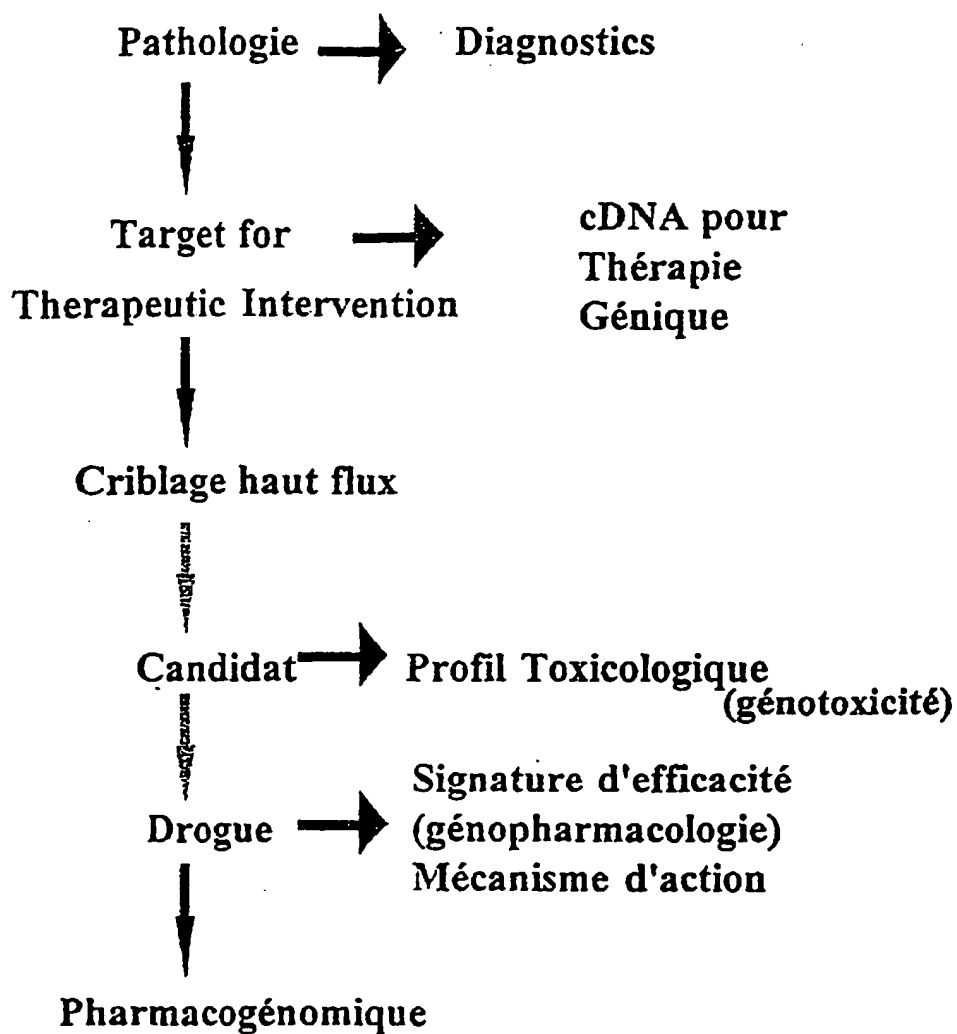
9/26

**A****B****Situations A et B****ARNm****1er brins d 'ADNc  
de A et de B****2ème brin d 'ADNc de A****Hybridations des 2 brins  
de A avec le 1er brin de B****Digestion par Sau3AI****Séparation des boucles engagées en duplex****Adjonction de linkers aux sites Sau3AI****Amplification par PCR puis clonage****FIGURE 6A**

10/26

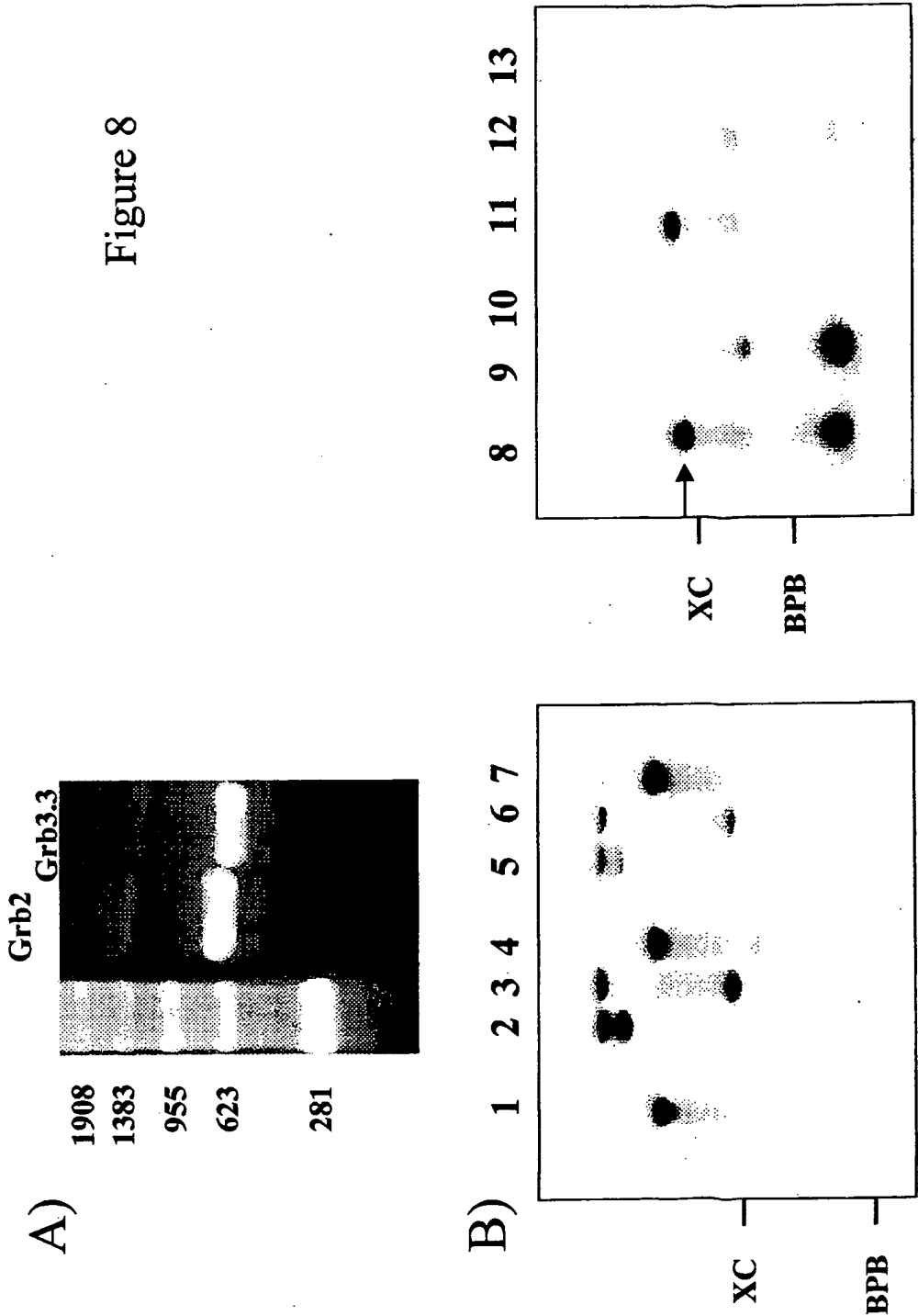
**FIGURE 6B**

11/26

**FIGURE 7**

12/26

Figure 8



13/26

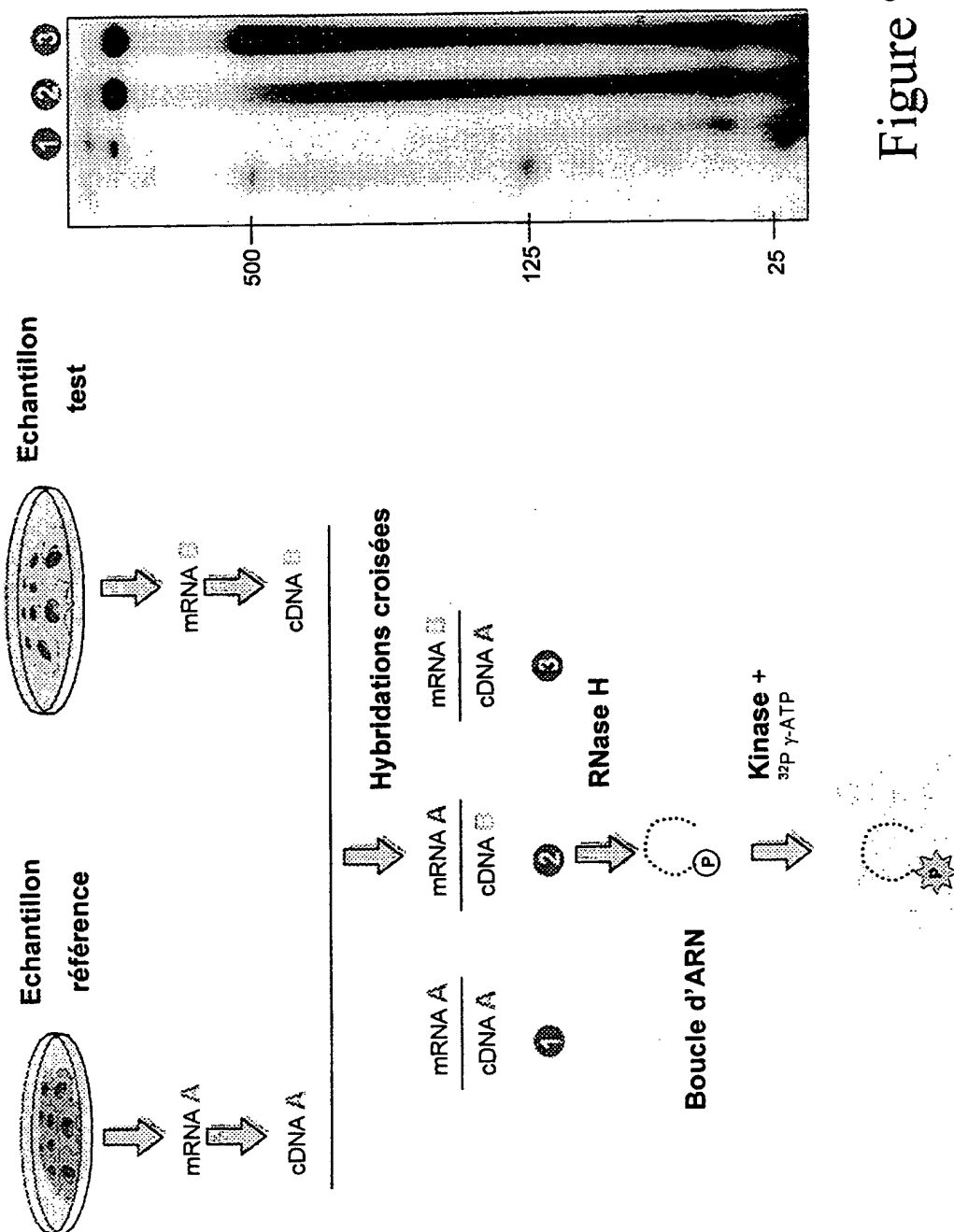


Figure 9



14/26

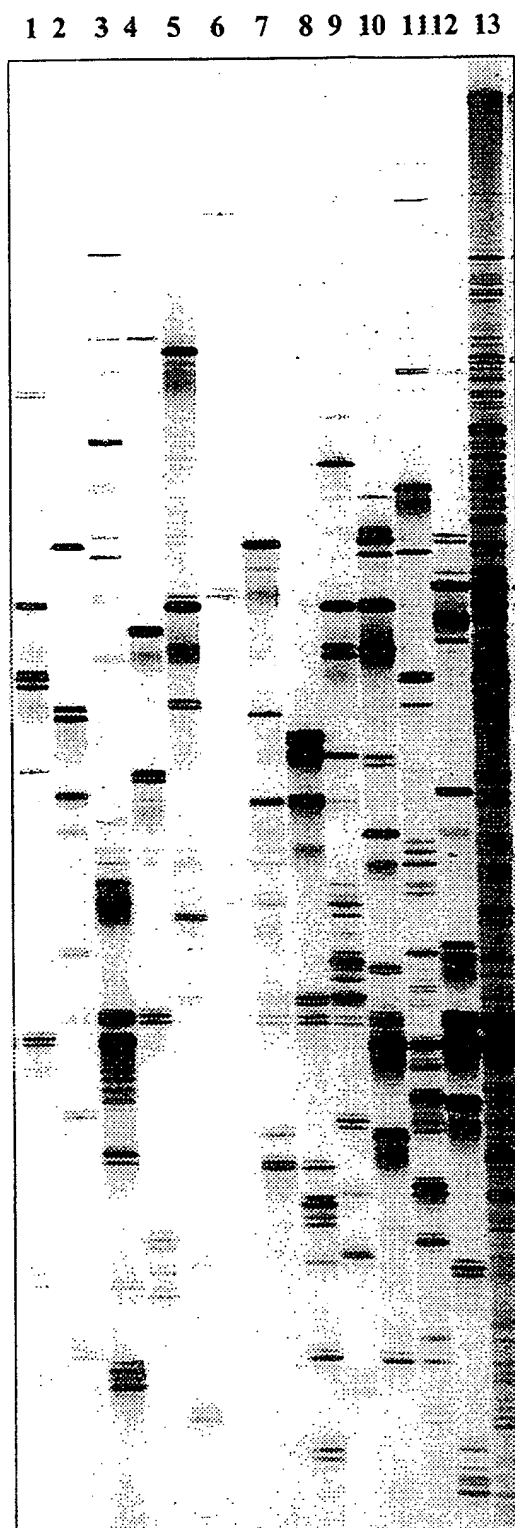
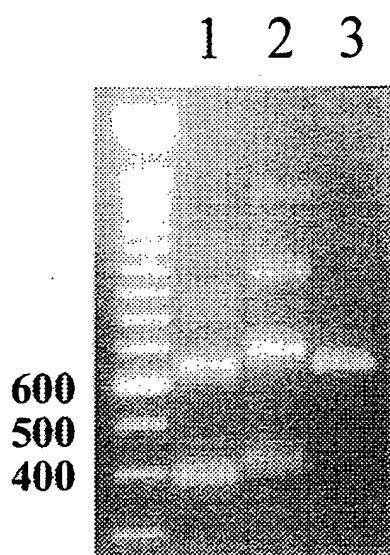


Figure 10

15/26

A)



B)

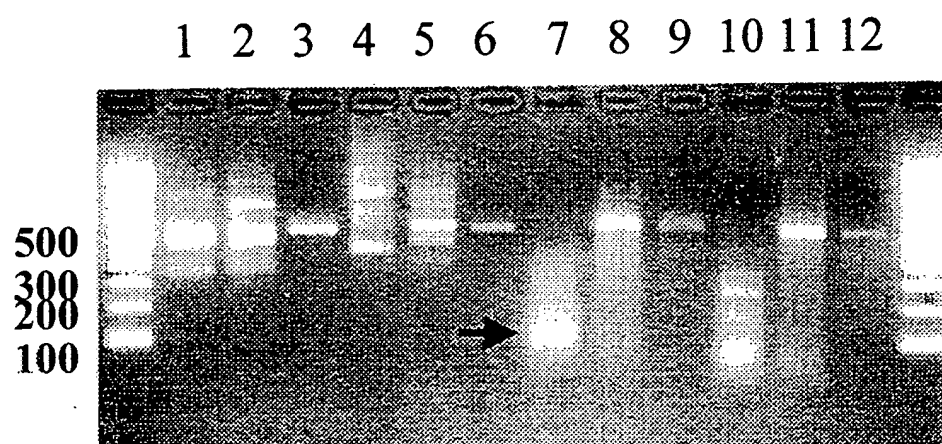


Figure 11

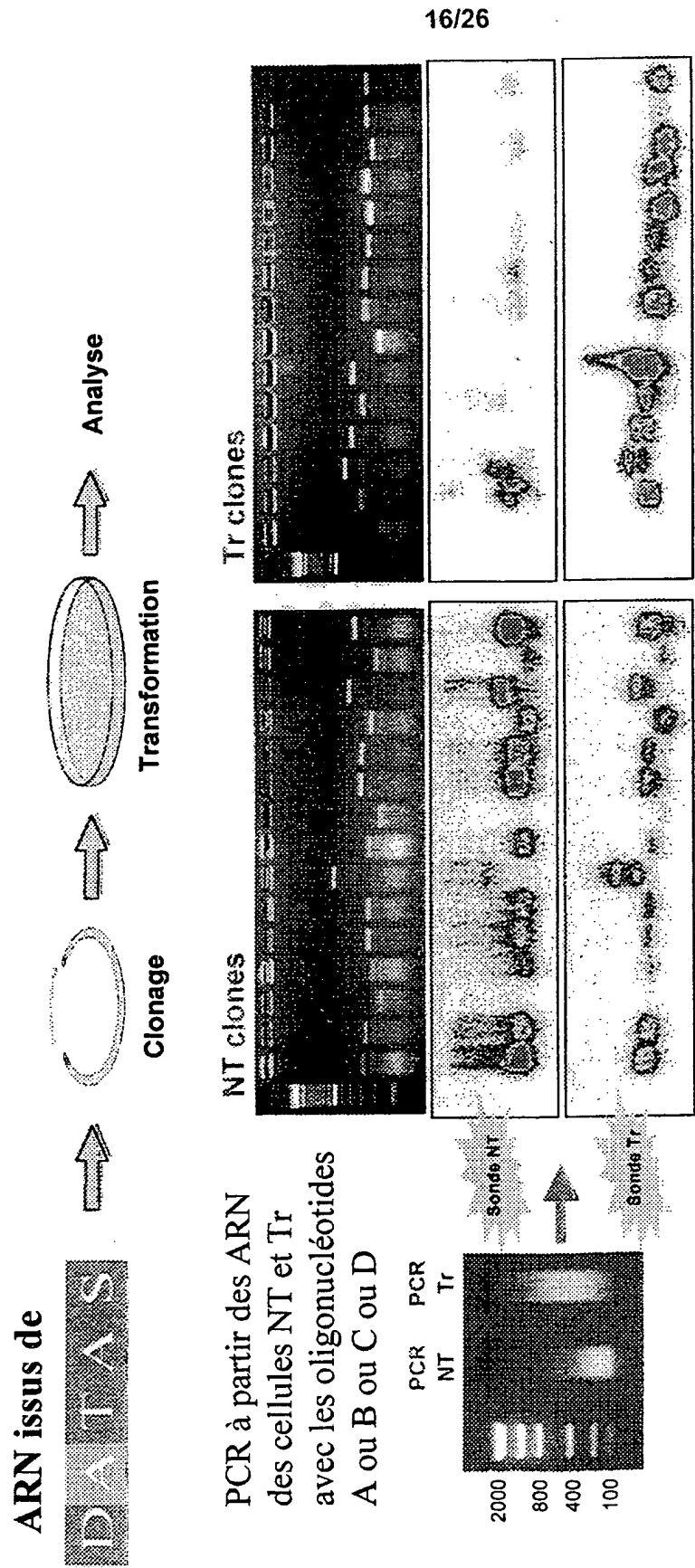
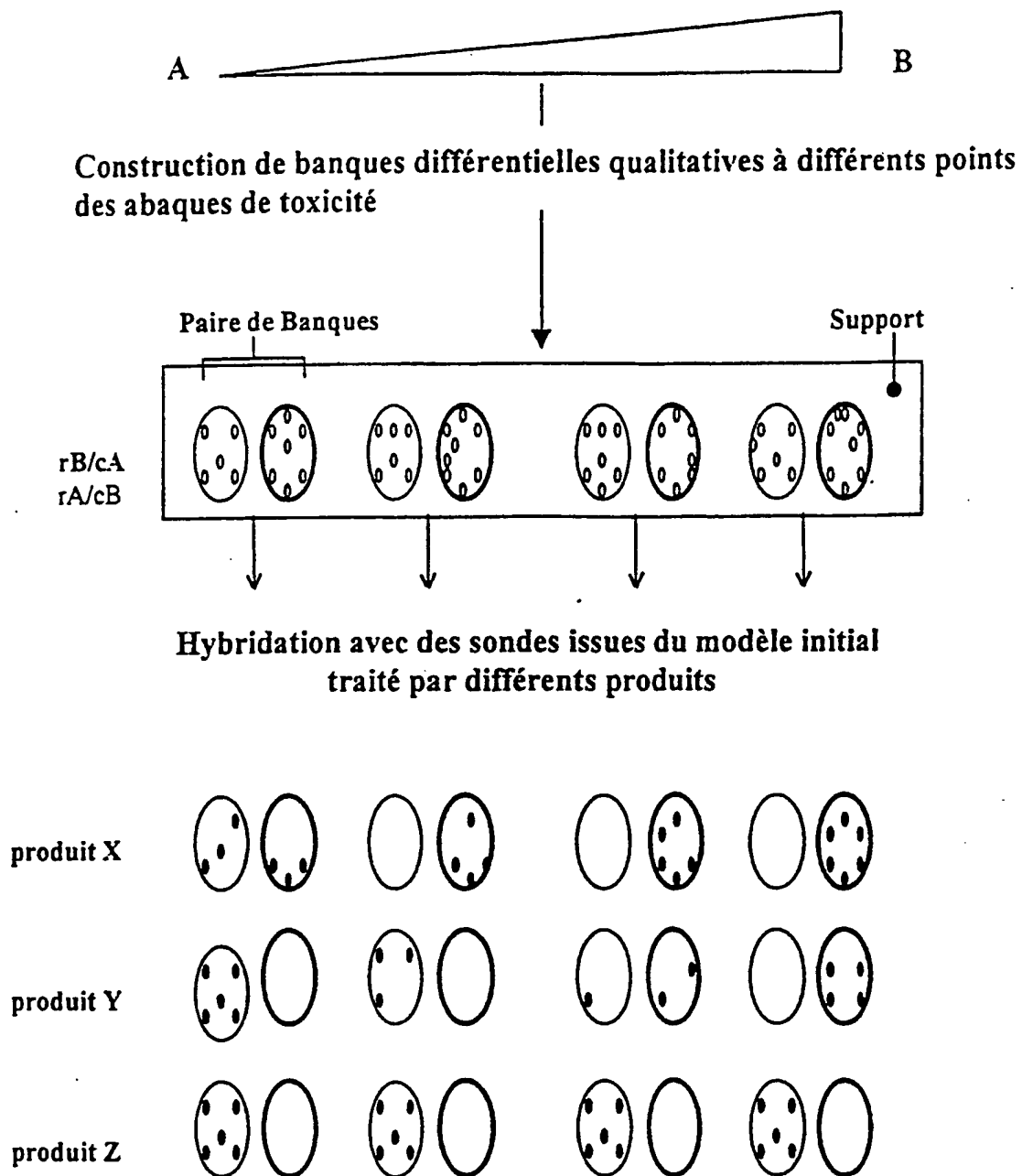
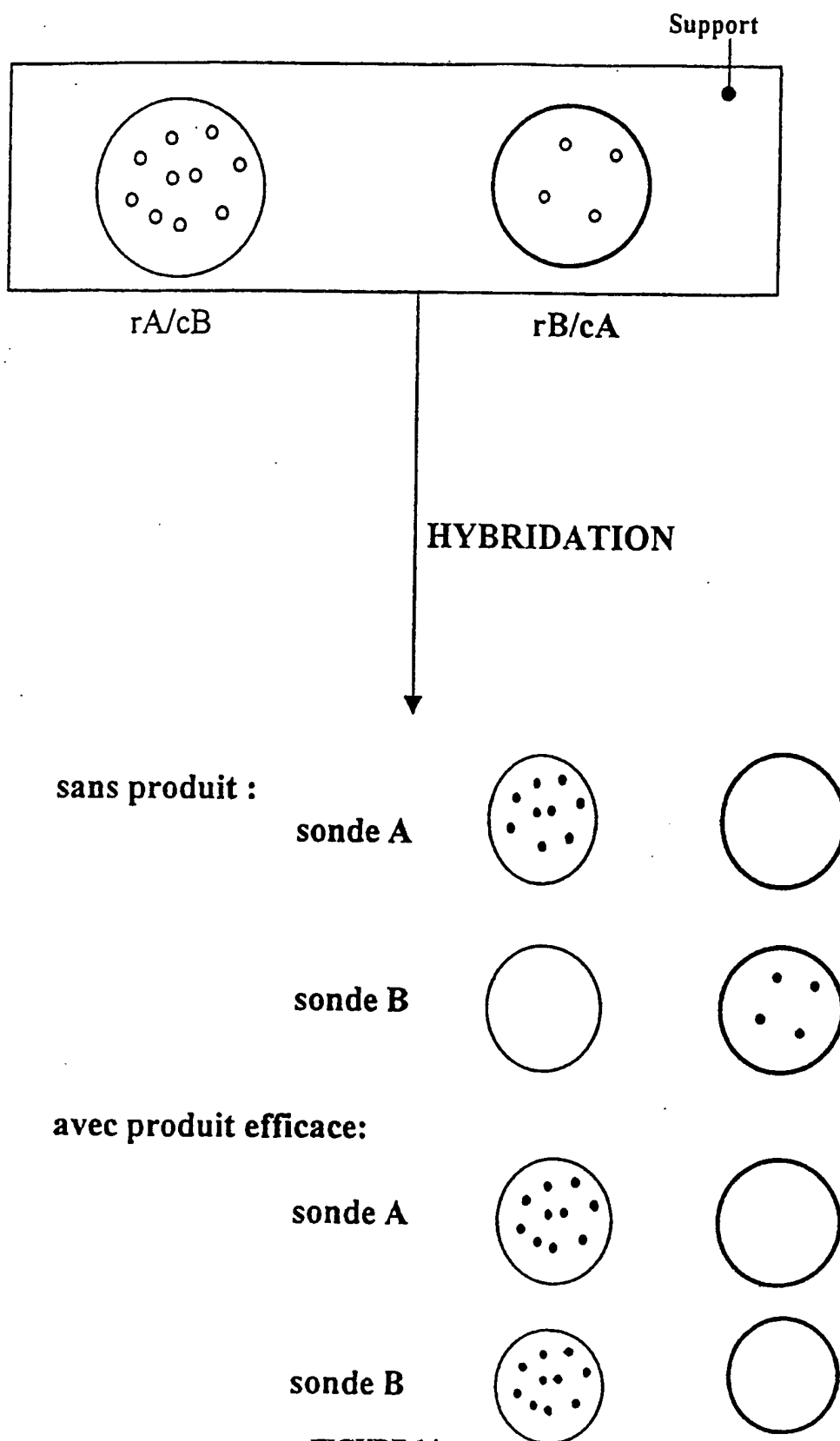


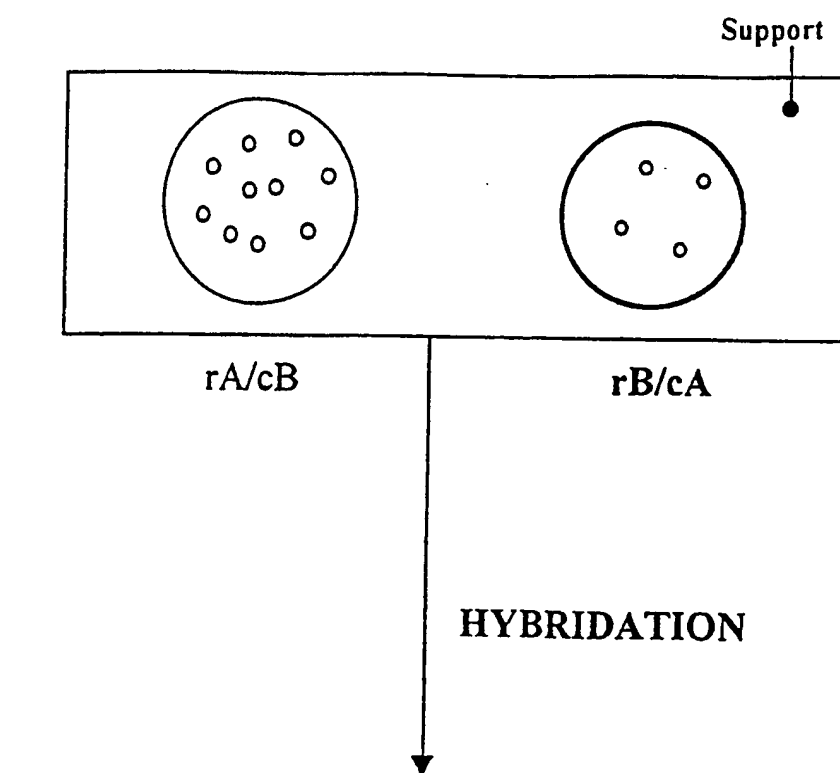
Figure 12

17/26

**FIGURE 13**

18/26

**FIGURE 14**



**biopsie de répondeurs:**



**biopsies de non répondeurs:**



**FIGURE 15**

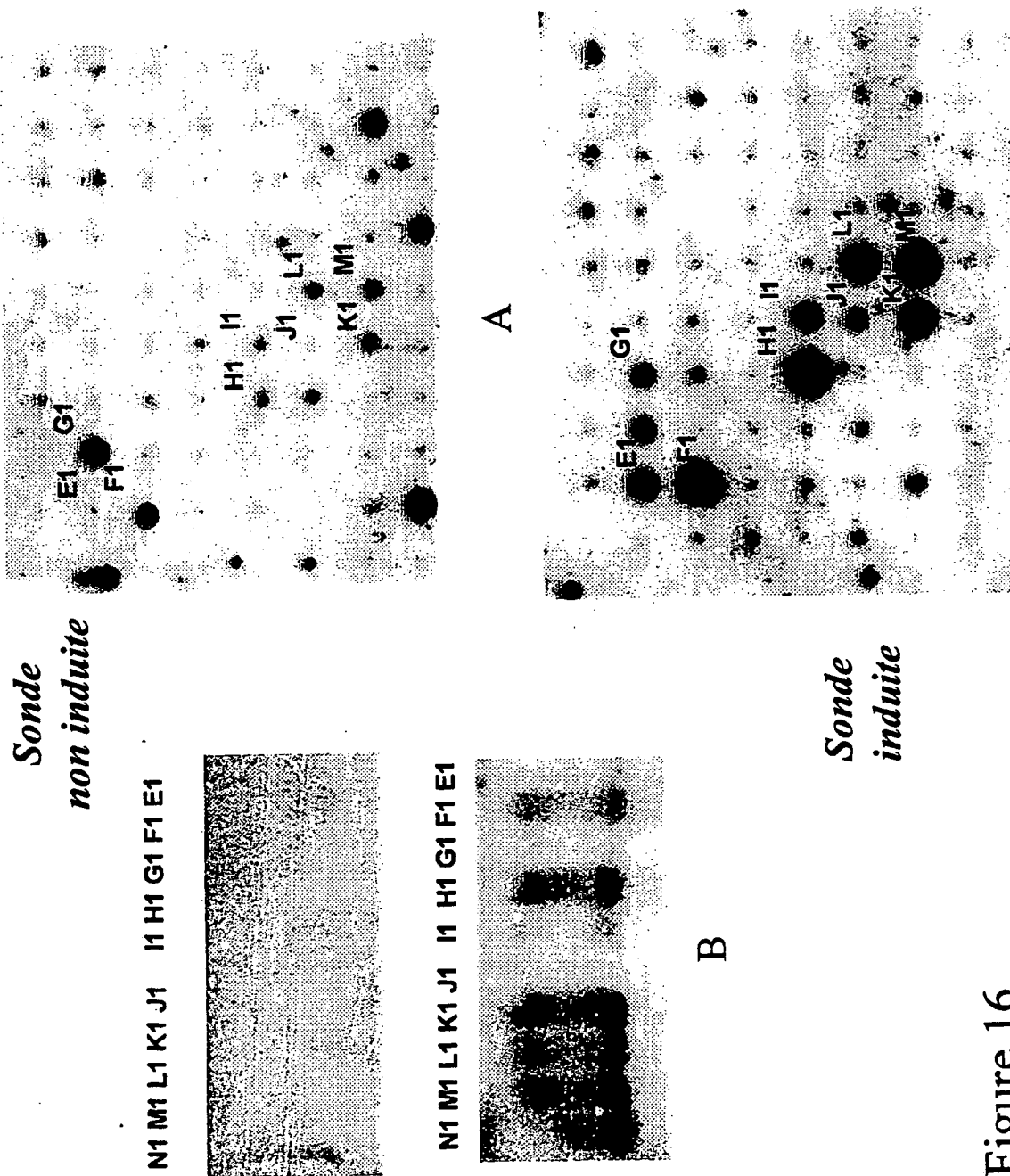


Figure 16

21/26

**SEQUENCE PEPTIDIQUE DE ΔSHC (SEQ ID NO: 9)**

1

MNKLSGGGGR RTRVEGGQLG GEEWTRHGSF VNKPTRGWLH PNDKVMGPGV  
 SYLVRYMGCV EVLQSMRALD FNTRTQVTRE AISLVCEAVP GAKGATRRRK  
 PCSRPLSSIL GRSNLKFAGM PITLTVSTSS LNLMAADCKQ IIANHHMQSI  
 SFASGGDPDT AEYVAYVAKD PVNQACHIL ECPEGLAQDV ISTIGQAFEL  
 RFKQYLRNPP KLVTPHDRMA GFDGSAWDEE EEEPPDHQYY NDFPGKEPPL  
 GGVDMLRE GAAPGAARPT APNAQTPSHL GATLPVGQPV GGDPEVRKQM  
 PPPPPCPGRE LFDDPSYVNV QNLDKARQAV GGAGPPNPAI NGSAPRDLFD  
 MKPFEDALRV PPPPQSVSMA EQLRGEFWFH GKLSRREAEA LLQLNGDFLV  
 RTKDHREFSV SHLISYHMDN HLPISAGSE LCLQQPVERKL

441

**SEQUENCE NUCLEIQUE DE ΔSHC (SEQ ID NO: 10)**

atgaacaagc	tgagtggagg	cggcgggccc	aggactcggg	tggaaggggg	50
ccagcttggg	ggcgaggagt	ggacccgcca	cgggagcttt	gtcaataagc	100
ccacgcgggg	ctggctgcat	cccaacgaca	aagtcattgg	acccgggggt	150
tcctacttgg	ttcgggtacat	gggttgtgtg	gaggtcctcc	agtcaatgcg	200
tgccctggac	ttcaacaccc	ggactcaggt	caccagggag	gccatcagtc	250
tggtgtgtga	ggctgtgccg	ggtgctaagg	gggcgacaag	gaggagaaag	300
ccctgtagcc	gcccgtcag	ctctatcctg	gggaggagta	acctgaaatt	350
tgctggaatg	ccaatcactc	tcaccgtctc	caccagcagc	ctcaacctca	400
tggccgcaga	ctgcaaacag	atcatcgcca	accaccacat	gcaatctatc	450
tcatttgcac	ccggcgggga	tccggacaca	gccgagtatg	tcgcctatgt	500
tgccaaagac	cctgtgaatc	agagagcctg	ccacattctg	gagtgtcccg	550
aagggttgc	ccaggatgtc	atcagcacca	ttggccaggc	cttcgagttg	600
cgttcaaac	aatacctcag	gaaccacccc	aaactgggtc	cccctcatga	650
caggatggct	ggctttgatg	gctcagcatg	ggatgaggag	gaggaagagc	700
cacctgacca	tcagtactat	aatgacttcc	cggggaagga	accccccttg	750
gggggggtgg	tagacatgag	gcttcgggaa	ggagccgctc	caggggctgc	800
tcgaccact	gcacccaatg	cccagacccc	cagccacttg	ggagctacat	850
tgcccttagg	acagcctgtt	gggggagatc	cagaagtccg	caaacagatg	900

**FIGURE 17A**



22/26

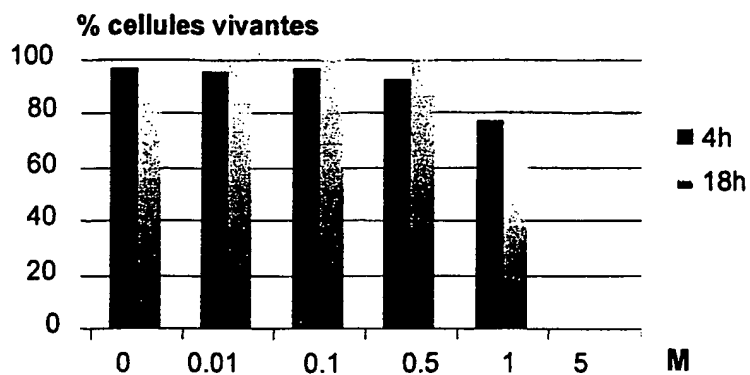
```
ccacctccac caccctgtcc aggcagagag ctttttgatg atccctccta 950
tgtcaacgtc cagaacctag acaaggcccg gcaagcagtg ggtgggtgctg 1000
ggccccccaa tctgtctatc aatggcagtg caccocggga cctgtttgac 1050
atgaagccct tcgaagatgc tcttcgggtg cctccacctc cccagtcggt 1100
gtccatggct gagcagctcc gaggggagcc ctggttccat gggaagctga 1150
gccggcggga ggctgaggca ctgctgcagc tcaatgggga cttcttggtt 1200
cggactaagg atcacgctt tgaaagtgtc agtcacctta tcagctacca 1250
catggacaat cacttgccca tcctctctgc gggcagcgaa ctgtgtctac 1300
agcaacctgt ggagcggaaa ctgtga 1326
```

**FIGURE 17B**

23/26

Bleu Trypan

HepG2 / Ethanol



Test MTT

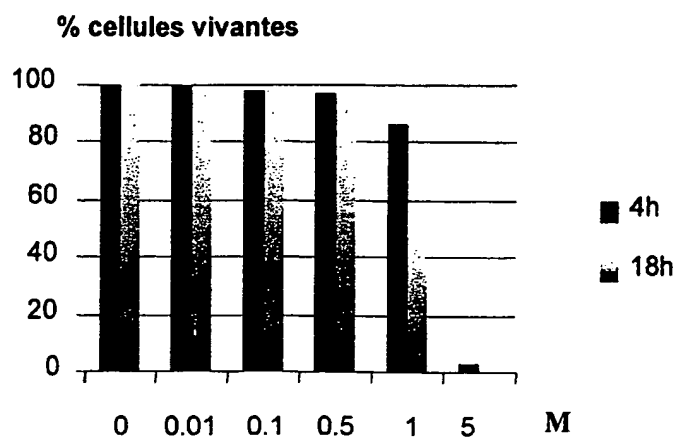
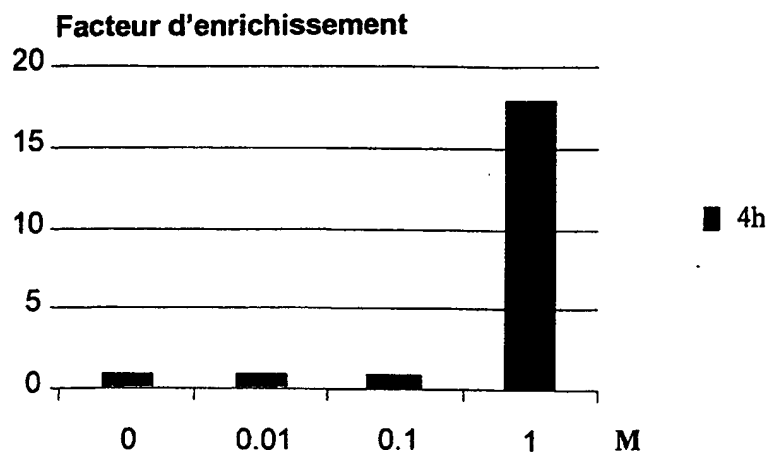


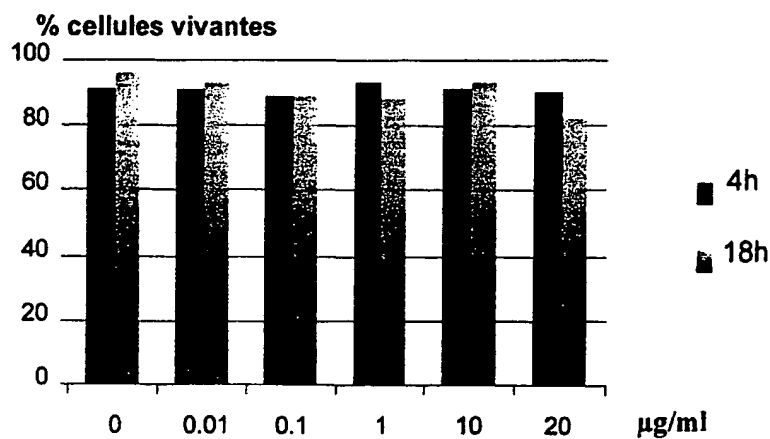
FIGURE 18A

Test ELISA de fragmentation de l'ADN

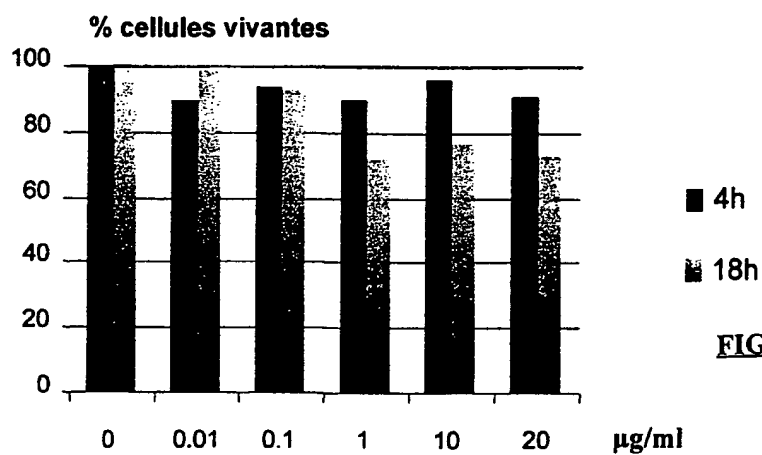


24/26

# **Bleu Trypan** HepG2 / Camptothecin

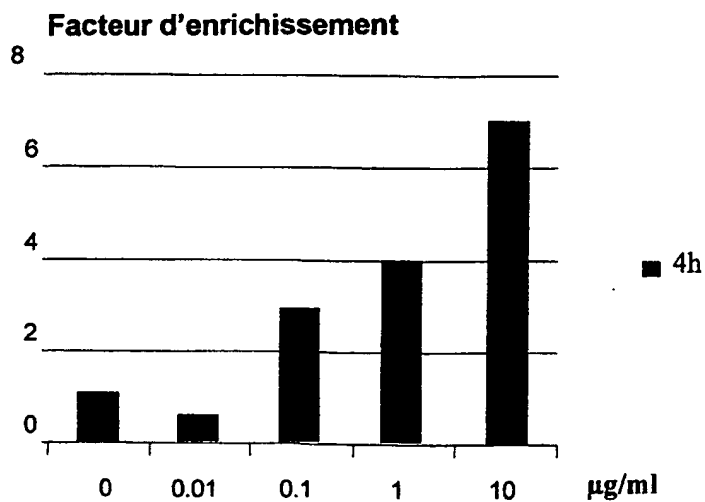


## **Test MTT**

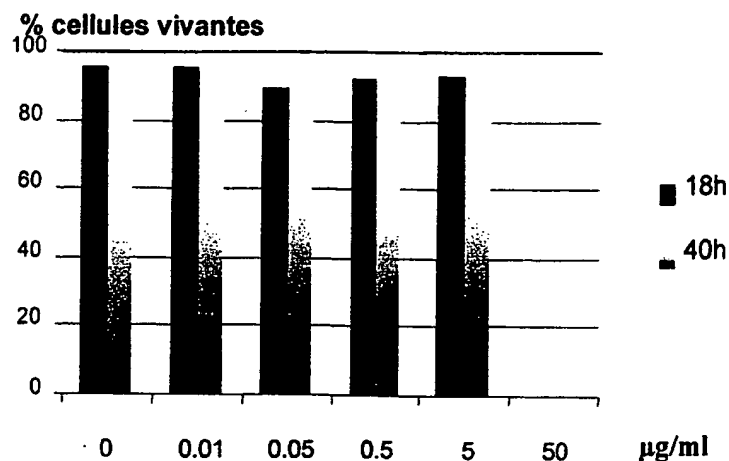
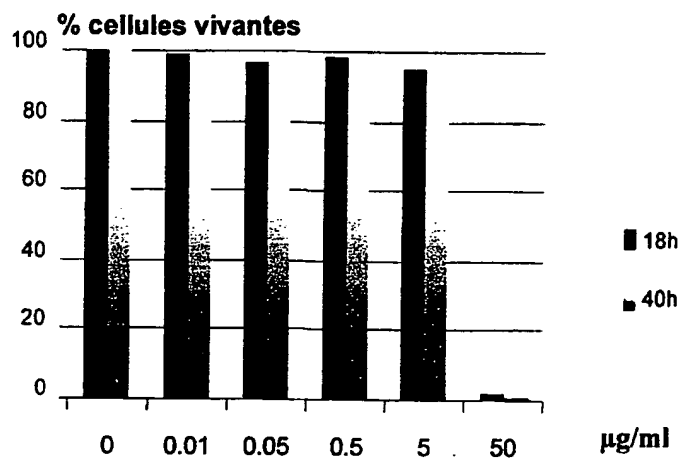
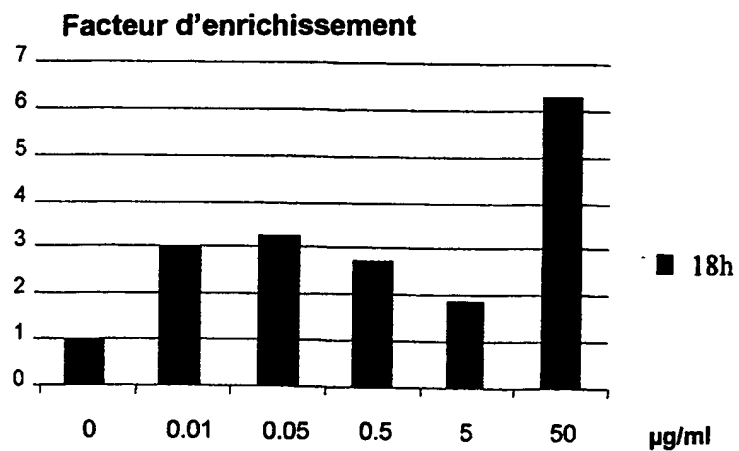


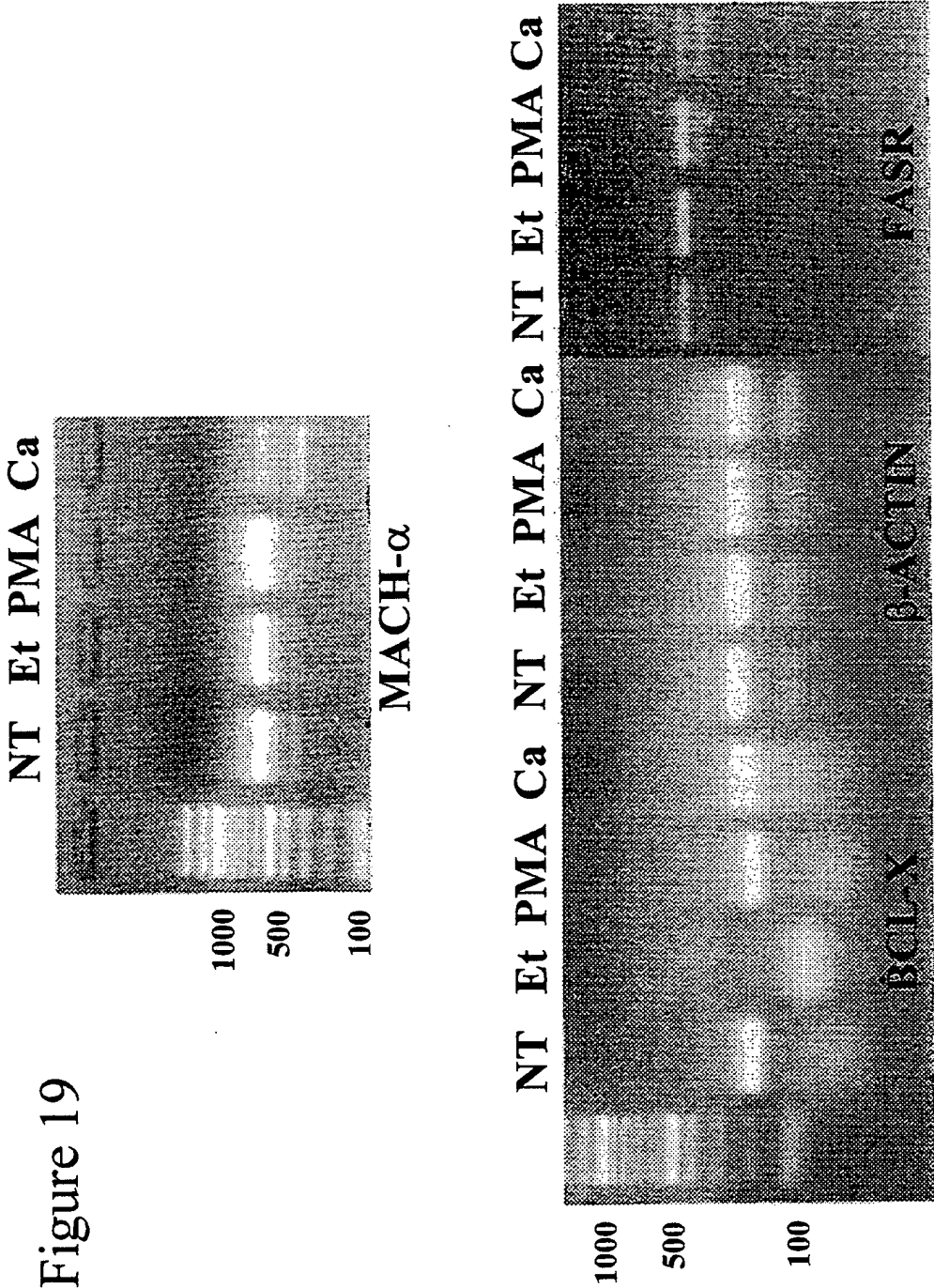
**FIGURE 18B**

## **Test ELISA de fragmentation de l'ADN**



25/26

**Bleu Trypan** **HepG2 / PMA****Test MTT****Test ELISA de fragmentation de l'ADN****FIGURE 18C**



## LISTE DE SEQUENCES

<110> EXONHIT THERAPEUTICS SA

<120> CRIBLAGE DIFFERENTIEL QUALITATIF

<130> B3898B - PB/KM

<140>

<141>

<150> 9802997

<151> 1998-03-11

<160> 16

<170> PatentIn Ver. 2.1

<210> 1

<211> 23

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: OLIGO

<400> 1

gagaagcgtt atnnnnnnna ggn

23

<210> 2

<211> 24

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: OLIGO

<400> 2

gagaagcgtt atnnnnnnnn tccc

24

<210> 3

<211> 23

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: OLIGO

<400> 3

gagaagcggtt atnnnnnnnn nnn

23

<210> 4

<211> 20

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: OLIGO

<400> 4

gagaagcggtt atnnnnncca

20

<210> 5

<211> 66

<212> ADN

<213> Homo sapiens

<400> 5

ccacacctgg ccagtatgtg ctcaactggct tgcagagtgg gcagccagcc taagcatttg 60  
cactgg 66

<210> 6

<211> 23

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: OLIGO

<400> 6

gggacctggtt tgacatgaag ccc

23

<210> 7

<211> 22

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: OLIGO

<400> 7

cagtttccgc tccacagggtt gc

22

<210> 8  
 <211> 96  
 <212> ADN  
 <213> Homo sapiens

<400> 8  
 gtacgggaga gcacgaccac acctggccag tatgtgctca ctggcttgca gaggggcag 60  
 cctaagcatt tgctactggt ggaccctgag ggtgtg 96

<210> 9  
 <211> 441  
 <212> PRT  
 <213> Homo sapiens

<400> 9  
 Met Asn Lys Leu Ser Gly Gly Gly Gly Arg Arg Thr Arg Val Glu Gly  
     1                    5                    10                    15  
 Gly Gln Leu Gly Gly Glu Glu Trp Thr Arg His Gly Ser Phe Val Asn  
                     20                    25                    30  
 Lys Pro Thr Arg Gly Trp Leu His Pro Asn Asp Lys Val Met Gly Pro  
             35                    40                    45  
 Gly Val Ser Tyr Leu Val Arg Tyr Met Gly Cys Val Glu Val Leu Gln  
     50                    55                    60  
 Ser Met Arg Ala Leu Asp Phe Asn Thr Arg Thr Gln Val Thr Arg Glu  
     65                    70                    75                    80  
 Ala Ile Ser Leu Val Cys Glu Ala Val Pro Gly Ala Lys Gly Ala Thr  
                     85                    90                    95  
 Arg Arg Arg Lys Pro Cys Ser Arg Pro Leu Ser Ser Ile Leu Gly Arg  
             100                    105                    110  
 Ser Asn Leu Lys Phe Ala Gly Met Pro Ile Thr Leu Thr Val Ser Thr  
     115                    120                    125  
 Ser Ser Leu Asn Leu Met Ala Ala Asp Cys Lys Gln Ile Ile Ala Asn  
     130                    135                    140  
 His His Met Gln Ser Ile Ser Phe Ala Ser Gly Gly Asp Pro Asp Thr  
     145                    150                    155                    160  
 Ala Glu Tyr Val Ala Tyr Val Ala Lys Asp Pro Val Asn Gln Arg Ala  
             165                    170                    175  
 Cys His Ile Leu Glu Cys Pro Glu Gly Leu Ala Gln Asp Val Ile Ser  
     180                    185                    190  
 Thr Ile Gly Gln Ala Phe Glu Leu Arg Phe Lys Gln Tyr Leu Arg Asn  
     195                    200                    205



Pro Pro Lys Leu Val Thr Pro His Asp Arg Met Ala Gly Phe Asp Gly  
210 215 220

Ser Ala Trp Asp Glu Glu Glu Glu Pro Pro Asp His Gln Tyr Tyr  
225 230 235 240

Asn Asp Phe Pro Gly Lys Glu Pro Pro Leu Gly Gly Val Val Asp Met  
245 250 255

Arg Leu Arg Glu Gly Ala Ala Pro Gly Ala Ala Arg Pro Thr Ala Pro  
260 265 270

Asn Ala Gln Thr Pro Ser His Leu Gly Ala Thr Leu Pro Val Gly Gln  
275 280 285

Pro Val Gly Gly Asp Pro Glu Val Arg Lys Gln Met Pro Pro Pro Pro  
290 295 300

Pro Cys Pro Gly Arg Glu Leu Phe Asp Asp Pro Ser Tyr Val Asn Val  
305 310 315 320

Gln Asn Leu Asp Lys Ala Arg Gln Ala Val Gly Gly Ala Gly Pro Pro  
325 330 335

Asn Pro Ala Ile Asn Gly Ser Ala Pro Arg Asp Leu Phe Asp Met Lys  
340 345 350

Pro Phe Glu Asp Ala Leu Arg Val Pro Pro Pro Pro Gln Ser Val Ser  
355 360 365

Met Ala Glu Gln Leu Arg Gly Glu Pro Trp Phe His Gly Lys Leu Ser  
370 375 380

Arg Arg Glu Ala Glu Ala Leu Leu Gln Leu Asn Gly Asp Phe Leu Val  
385 390 395 400

Arg Thr Lys Asp His Arg Phe Glu Ser Val Ser His Leu Ile Ser Tyr  
405 410 415

His Met Asp Asn His Leu Pro Ile Ile Ser Ala Gly Ser Glu Leu Cys  
420 425 430

Leu Gln Gln Pro Val Glu Arg Lys Leu  
435 440

<210> 10

<211> 1326

<212> ADN

<213> Homo sapiens

<400> 10

atgaacaagc tgagtggagg cggcgggagc aggactcggg tggaaggggg ccagcttggg 60  
ggcgaggagt ggaccgcga cgggagcttt gtcaataagc ccacgcgggg ctggctgcat 120

```

cccaacgaca aagtcacatggg acccggggtt tcctacttgg ttcggtacat ggggtgtgtg 180
gaggctcctcc agtcaatgcg tgccctggac ttcaacaccc ggactcaggt caccagggag 240
gccatcagtc tgggtgtgtga ggctgtgccg ggtgctaagg gggcgacaag gaggagaaag 300
ccctgtagcc gcccgtcag ctctatcctg gggaggagta acctgaaatt tgctggaatg 360
ccaatcactc tcaccgtctc caccagcagc ctcaacctca tggccgcaga ctgcaaacag 420
atcatcgcca accaccacat gcaatctatc tcatttgcac ccggcgggga tccggacaca 480
gccgagtatg tcgcttatgt tgccaaagac cctgtgaatc agagagcctg ccacattctg 540
gagtgtcccg aagggttgc ccaggatgtc atcagcacca ttggccaggc cttcgagttg 600
cgcttcaaac aatacctcag gaacccaccc aaactggtca cccctcatga caggatggct 660
ggctttgatg gctcagcatg ggatgaggag gaggaagagc cacctgacca tcagtactat 720
aatgacttcc cggggaagga accccccttg gggggggttg tagacatgag gcttcgggaa 780
ggagccgctc caggggctgc tcgaccact gcacccaatg ccagacccc cagccacttg 840
ggagctacat tgctgtagg acagcctgtt gggggagatc cagaagtccg caaacagatg 900
ccacctccac caccctgtcc aggcagagag ctttttgatg atccctccta tgtcaacgtc 960
cagaacctag acaaggcccg gcaagcagtg ggtggtgctg ggccccccaa tcctgctatc 1020
aatggcagtg caccgggga cctgtttgac atgaagccct tcgaagatgc tcttcgggtg 1080
cctccacctc ccagtcggt gtccatggct gagcagctcc gaggggagcc ctggttccat 1140
gggaagctga gccggcgga ggctgaggca ctgctgcagc tcaatgggga cttcttggtt 1200
cggactaagg atcaccgctt tgaaagtgtc agtcaccta tcagctacca catggacaat 1260
cacttgccca tcctctctgc gggcagcgaa ctgtgtctac agcaacctgt ggagcggaaa 1320
ctgtga 1326

```

&lt;210&gt; 11

&lt;211&gt; 19

&lt;212&gt; ADN

&lt;213&gt; Séquence artificielle

&lt;220&gt;

&lt;223&gt; Description de la séquence artificielle: OLIGO

&lt;400&gt; 11

tgcccaaatac aacaagagc

19

&lt;210&gt; 12

&lt;211&gt; 19

&lt;212&gt; ADN

&lt;213&gt; Séquence artificielle

&lt;220&gt;

&lt;223&gt; Description de la séquence artificielle: OLIGO

&lt;400&gt; 12

cccctgacaa gcctgaata

19

&lt;210&gt; 13

&lt;211&gt; 24

&lt;212&gt; ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: OLIGO

<400> 13

atgtctcaga gcaaccggga gctg

24

<210> 14

<211> 24

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: OLIGO

<400> 14

gtggctccat tcaccgcgga gctg

24

<210> 15

<211> 19

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: OLIGO

<400> 15

tgccaagaag ggaaggagt

19

<210> 16

<211> 20

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: OLIGO

<400> 16

tgtcatgact ccagcaatag

20

# INTERNATIONAL SEARCH REPORT

Int. l. Application No

PCT/FR 99/00547

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 6 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	COOPER D L ET AL: "GENE THERAPY ADVANCES: UTILIZATION OF ALTERNATIVE SPLICING AS A CONTROL ELEMENT IN THE CHIMERIC ENZYME/PRODRUG THERAPY (CEPT) APPROACH TO PRIMARY AND METASTATIC TUMORS" JOURNAL OF CLINICAL LIGAND ASSAY, vol. 19, no. 1, 1996, pages 80-84, XP002038774 see the whole document ---	1,5-8, 13,14, 26-28, 31,32, 38-42, 44-47
X	WO 95 27052 A (UNIV MARYLAND) 12 October 1995	14,15, 26,27
Y	see abstract; claims 1-16 ---	38,39,41
	-/--	

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

\* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

1 July 1999

Date of mailing of the international search report

15/07/1999

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Knehr, M

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/FR 99/00547

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 96 26272 A (HESTON WARREN D W ;SLOAN KETTERING INST CANCER (US); ISRAELI RON S) 29 August 1996	1,5-7, 10, 13-15, 26-28, 44-46
Y	see abstract; claims 1-20; example 6 ---	11,12,39
X	WO 98 02576 A (UNIV MASSACHUSETTS) 22 January 1998	1,5-7, 10, 13-15, 26-28,44
Y	see abstract; claims 1,4,7,10,11,14 ---	2-4,11, 12, 16-22, 45,46
X	FR 2 664 287 A (INST NAT SANTE RECH MED) 10 January 1992	1,5,6, 10, 13-15, 44,47
Y	see abstract; claims 1-15 ---	2-4,8
X	EP 0 330 781 A (TOA NENRYO KOGYO KK) 6 September 1989 see abstract see page 11, line 23 - line 25 ---	55
X	EP 0 709 397 A (HEALTH RESEARCH INC) 1 May 1996 see abstract; claims 1-10 ---	47
Y	ALPHEY L: "PCR-based method isolation of full-length clones and splice variants from cDNA libraries" BIOTECHNIQUES, vol. 22, 1997, pages 481-486, XP002087172 see the whole document ---	16-28, 33-35, 38-43
Y	ARDLEY H C ET AL.: "Rapid isolation of genomic clones for individual members of human multigene families: Identification and localisation of UBE2L4, a novel member of a ubiquitin conjugating enzyme dispersed gene family" CYTOGENETICS AND CELL GENETICS, vol. 79, 1997, pages 188-192, XP002087173 see abstract see page 188, column 1, paragraph 1 - page 190, column 1, paragraph 1; figure 1 ---	16-28, 33-35, 38-43
	-/--	

# INTERNATIONAL SEARCH REPORT

In. .ational Application No

PCT/FR 99/00547

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	BONASS W A ET AL.: "The rat amelogenin gene - some aspects of evolution and expression" ADVANCES IN DENTAL RESEARCH, vol. 10, no. 2, 1996, pages 182-186, XP002087174 see the whole document ---	16-28, 33-35, 38-43
Y	WO 94 12631 A (ISIS INNOVATION ;TARIN DAVID (GB); MATSUMURA YASUHIRO (GB)) 9 June 1994  see the whole document ---	1,5-8, 13-15, 26-28, 44-47
Y	WO 97 46679 A (ST JUDE CHILDRENS RES HOSPITAL) 11 December 1997  see abstract; claims 1-32 ---	1,5-8, 10-15, 26-32, 39,44,47
Y	EP 0 806 478 A (HEALTH RESEARCH INC) 12 November 1997  see the whole document ---	1,7,30, 42,43, 45,46
Y	WO 96 30512 A (RHONE POULENC RORER SA ;BRACCO LAURENT (FR); SCHWEIGHOFFER FABIEN) 3 October 1996  see the whole document ---	1,11-16, 20,23, 26-32, 38-41,44
Y	US 5 679 541 A (LEISERSON WILLIAM M ET AL) 21 October 1997  see abstract see column 21, line 9 - line 36 ---	1,5,6,9, 16, 20-22,31
Y	WO 97 04092 A (RHONE POULENC RORER SA ;CONSEILLER EMMANUEL (FR); BRACCO LAURENT ()) 6 February 1997 see abstract; claims 1-53 ---	1,5-10, 30,42, 43,45,46
Y	MILLER R D AND RIBLET R: "Improved phenol emulsion DNA reassociation technique (PERT) using thermal cycling" NUCLEIC ACIDS RESEARCH, vol. 23, no. 12, 1995, pages 2339-2340, XP002087175 cited in the application see the whole document ---	1,10
A	EP 0 791 660 A (SMITHKLINE BEECHAM CORP ;UNIV WASHINGTON (US); UNIV SOUTH FLORIDA) 27 August 1997 see the whole document ---	
	---	

# INTERNATIONAL SEARCH REPORT

In. ational Application No

PCT/FR 99/00547

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>DIATCHENKO L ET AL: "SUPPRESSION SUBTRACTIVE HYBRIDIZATION: A METHOD FOR GENERATING DIFFERENTIALLY REGULATED OR TISSUE-SPECIFIC CDNA PROBES AND LIBRARIES" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 93, 1 June 1996, pages 6025-6030, XP002911922 cited in the application see the whole document</p> <p style="text-align: center;">---</p>	
A	<p>PERRET E ET AL: "Improved differential screening approach to analyse transcriptional variations in organized cDNA libraries" GENE: AN INTERNATIONAL JOURNAL ON GENES AND GENOMES, vol. 208, no. 2, 22 February 1998, page 103-115 XP004114934 see the whole document</p> <p style="text-align: center;">---</p>	
A	<p>PELICCI G ET AL: "A NOVEL TRANSFORMING PROTEIN (SHC) WITH AN SH2 DOMAIN IS IMPLICATED IN MITOGENIC SIGNAL TRANSDUCTION" CELL, vol. 70, no. 1, 10 July 1992, pages 93-104, XP002001630 cited in the application see abstract; figure 1</p> <p style="text-align: center;">---</p>	
A	<p>HARUN R B ET AL.: "Characterization of human SHC p66 cDNA and its processed pseudogene mapping to Xq12-q13.1" GENOMICS, vol. 42, 1997, pages 349-352, XP002107843 see the whole document</p> <p style="text-align: center;">-----</p>	

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/FR 99/00547

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
  
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see supplementary sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☒ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
  
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.  
☐ No protest accompanied the payment of additional search fees.



The International Searching Authority found several (groups of) inventions in the international application, namely:

1. Claims 1-47, 52-57 (wholly)

Methods for identifying and/or cloning nucleic acid regions representing qualitative differences between two (patho) biological samples, oligonucleotides suitable for cloning said nucleic acids, nucleic acids identified by said methods, compositions and banks comprising said nucleic acids, kits comprising said banks, use of said nucleic acids, compositions or banks, a method for identifying and/or producing proteins related to a pathological condition, a protein identified or produced by said method, an antibody directed against said protein, a method for identifying and/or cloning tumour suppressor genes, a method for determining the toxicity of a compound, a method for determining the therapeutic efficacy of a compound, and a method for determining a patient's response to a compound.

2. Claims: 48-51 (wholly)

A protein of sequences SEQ ID NO:9, a nucleic probe or antibody for identifying said protein, a screening method, and a vector comprising a sequence coding for said protein.

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/FR 99/00547

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9527052 A	12-10-1995	AU 2237395 A CA 2186854 A EP 0775202 A JP 9511400 T ZA 9502639 A	23-10-1995 12-10-1995 28-05-1997 18-11-1997 02-01-1996
WO 9626272 A	29-08-1996	AU 5172596 A CA 2212846 A EP 0812356 A	11-09-1996 29-08-1996 17-12-1997
WO 9802576 A	22-01-1998	AU 3494297 A	09-02-1998
FR 2664287 A	10-01-1992	NONE	
EP 0330781 A	06-09-1989	JP 1218590 A CA 1313151 A DE 3855076 D DE 3855076 T	31-08-1989 26-01-1993 11-04-1996 17-10-1996
EP 0709397 A	01-05-1996	US 5688918 A CA 2150994 A JP 8081500 A US 5726024 A	18-11-1997 15-12-1995 26-03-1996 10-03-1998
WO 9412631 A	09-06-1994	AT 136940 T CA 2139410 A CA 2149635 A DE 69302276 D DE 69302276 T EP 0651822 A EP 0672130 A WO 9402633 A JP 2800850 B JP 8500731 T JP 8506801 T US 5830646 A US 5879898 A	15-05-1996 03-02-1994 09-06-1994 23-05-1996 19-09-1996 10-05-1995 20-09-1995 03-02-1994 21-09-1998 30-01-1996 23-07-1996 03-11-1998 09-03-1999
WO 9746679 A	11-12-1997	AU 3300397 A	05-01-1998
EP 0806478 A	12-11-1997	US 5747650 A US 5726024 A	05-05-1998 10-03-1998
WO 9630512 A	03-10-1996	FR 2732348 A AU 5402096 A BR 9607928 A CA 2214451 A CZ 9703080 A EP 0817845 A HU 9801221 A JP 11503011 T NO 974449 A SK 131197 A ZA 9602506 A	04-10-1996 16-10-1996 09-06-1998 03-10-1996 14-01-1998 14-01-1998 28-08-1998 23-03-1999 26-09-1997 06-05-1998 01-10-1996
US 5679541 A	21-10-1997	NONE	
WO 9704092 A	06-02-1997	FR 2736915 A	24-01-1997

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/FR 99/00547

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9704092 A		AU 6618696 A	18-02-1997
		CA 2224468 A	06-02-1997
		CZ 9800144 A	15-04-1998
		EP 0839194 A	06-05-1998
		NO 980203 A	10-03-1998
		SK 6398 A	09-09-1998
EP 0791660 A	27-08-1997	JP 10000100 A	06-01-1998

# RAPPORT DE RECHERCHE INTERNATIONALE

Document International No

PCT/FR 99/00547

## A. CLASSEMENT DE L'OBJET DE LA DEMANDE

CIB 6 C1201/68

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

## B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)

CIB 6 C120

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si réalisable, termes de recherche utilisés)

## C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie *	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	COOPER D L ET AL: "GENE THERAPY ADVANCES: UTILIZATION OF ALTERNATIVE SPLICING AS A CONTROL ELEMENT IN THE CHIMERIC ENZYME/PRODRUG THERAPY (CEPT) APPROACH TO PRIMARY AND METASTATIC TUMORS" JOURNAL OF CLINICAL LIGAND ASSAY, vol. 19, no. 1, 1996, pages 80-84, XP002038774 voir le document en entier ---	1,5-8, 13,14, 26-28, 31,32, 38-42, 44-47
X	WO 95 27052 A (UNIV MARYLAND) 12 octobre 1995	14,15, 26,27
Y	voir abrégé; revendications 1-16 ---	38,39,41
-/--		

☒ Voir la suite du cadre C pour la fin de la liste des documents

☒ Les documents de familles de brevets sont indiqués en annexe

### \* Catégories spéciales de documents cités:

"A" document définissant l'état général de la technique, non considéré comme particulièrement pertinent

"E" document antérieur, mais publié à la date de dépôt international ou après cette date

"L" document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)

"O" document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens

"P" document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

"T" document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention

"X" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément

"Y" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier

"Z" document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

1 juillet 1999

Date d'expédition du présent rapport de recherche internationale

15/07/1999

Nom et adresse postale de l'administration chargée de la recherche internationale

Office Européen des Brevets, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Fonctionnaire autorisé

Knehr, M

# RAPPORT DE RECHERCHE INTERNATIONALE

De. de Internationale No

PCT/FR 99/00547

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS		
Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	WO 96 26272 A (HESTON WARREN D W ;SLOAN KETTERING INST CANCER (US); ISRAELI RON S) 29 août 1996	1,5-7, 10, 13-15, 26-28, 44-46
Y	voir abrégé; revendications 1-20; exemple 6 ---	11,12,39
X	WO 98 02576 A (UNIV MASSACHUSETTS) 22 janvier 1998	1,5-7, 10, 13-15, 26-28,44
Y	voir abrégé; revendications 1,4,7,10,11,14 ---	2-4,11, 12, 16-22, 45,46
X	FR 2 664 287 A (INST NAT SANTE RECH MED) 10 janvier 1992	1,5,6, 10, 13-15, 44,47
Y	voir abrégé; revendications 1-15 ---	2-4,8
X	EP 0 330 781 A (TOA NENRYO KOGYO KK) 6 septembre 1989 voir abrégé voir page 11, ligne 23 - ligne 25 ---	55
X	EP 0 709 397 A (HEALTH RESEARCH INC) 1 mai 1996 voir abrégé; revendications 1-10 ---	47
Y	ALPHEY L: "PCR-based method isolation of full-length clones and splice variants from cDNA libraries" BIOTECHNIQUES, vol. 22, 1997, pages 481-486, XP002087172 voir le document en entier ---	16-28, 33-35, 38-43
Y	ARDLEY H C ET AL.: "Rapid isolation of genomic clones for individual members of human multigene families: Identification and localisation of UBE2L4, a novel member of a ubiquitin conjugating enzyme dispersed gene family" CYTOGENETICS AND CELL GENETICS, vol. 79, 1997, pages 188-192, XP002087173 voir abrégé voir page 188, colonne 1, alinéa 1 - page 190, colonne 1, alinéa 1; figure 1 ---	16-28, 33-35, 38-43
	---	
	-/--	

# RAPPORT DE RECHERCHE INTERNATIONALE

D n de Internationale No

PCT/FR 99/00547

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS		
Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
Y	BONASS W A ET AL.: "The rat amelogenin gene - some aspects of evolution and expression" ADVANCES IN DENTAL RESEARCH, vol. 10, no. 2, 1996, pages 182-186, XP002087174 voir le document en entier ---	16-28, 33-35, 38-43
Y	WO 94 12631 A (ISIS INNOVATION ;TARIN DAVID (GB); MATSUMURA YASUHIRO (GB)) 9 juin 1994  voir le document en entier ---	1,5-8, 13-15, 26-28, 44-47
Y	WO 97 46679 A (ST JUDE CHILDRENS RES HOSPITAL) 11 décembre 1997  voir abrégé; revendications 1-32 ---	1,5-8, 10-15, 26-32, 39,44,47
Y	EP 0 806 478 A (HEALTH RESEARCH INC) 12 novembre 1997  voir le document en entier ---	1,7,30, 42,43, 45,46
Y	WO 96 30512 A (RHONE POULENC RORER SA ;BRACCO LAURENT (FR); SCHWEIGHOFFER FABIEN) 3 octobre 1996  voir le document en entier ---	1,11-16, 20,23, 26-32, 38-41,44
Y	US 5 679 541 A (LEISERSON WILLIAM M ET AL) 21 octobre 1997  voir abrégé voir colonne 21, ligne 9 - ligne 36 ---	1,5,6,9, 16, 20-22,31
Y	WO 97 04092 A (RHONE POULENC RORER SA ;CONSEILLER EMMANUEL (FR); BRACCO LAURENT ( ) 6 février 1997 voir abrégé; revendications 1-53 ---	1,5-10, 30,42, 43,45,46
Y	MILLER R D AND RIBLET R: "Improved phenol emulsion DNA reassociation technique (PERT) using thermal cycling" NUCLEIC ACIDS RESEARCH, vol. 23, no. 12, 1995, pages 2339-2340, XP002087175 cité dans la demande voir le document en entier ---	1,10
A	EP 0 791 660 A (SMITHKLINE BEECHAM CORP ;UNIV WASHINGTON (US); UNIV SOUTH FLORIDA) 27 août 1997 voir le document en entier ---	

# RAPPORT DE RECHERCHE INTERNATIONALE

Document Internationale No

PCT/FR 99/00547

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS		
Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
A	<p>DIATCHENKO L ET AL: "SUPPRESSION SUBTRACTIVE HYBRIDIZATION: A METHOD FOR GENERATING DIFFERENTIALLY REGULATED OR TISSUE-SPECIFIC CDNA PROBES AND LIBRARIES" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 93, 1 juin 1996, pages 6025-6030, XP002911922 cité dans la demande voir le document en entier ---</p>	
A	<p>PERRET E ET AL: "Improved differential screening approach to analyse transcriptional variations in organized cDNA libraries" GENE: AN INTERNATIONAL JOURNAL ON GENES AND GENOMES, vol. 208, no. 2, 22 février 1998, page 103-115 XP004114934 voir le document en entier ---</p>	
A	<p>PELICCI G ET AL: "A NOVEL TRANSFORMING PROTEIN (SHC) WITH AN SH2 DOMAIN IS IMPLICATED IN MITOGENIC SIGNAL TRANSDUCTION" CELL, vol. 70, no. 1, 10 juillet 1992, pages 93-104, XP002001630 cité dans la demande voir abrégé; figure 1 ---</p>	
A	<p>HARUN R B ET AL.: "Characterization of human SHC p66 cDNA and its processed pseudogene mapping to Xq12-q13.1" GENOMICS, vol. 42, 1997, pages 349-352, XP002107843 voir le document en entier -----</p>	

# RAPPORT DE RECHERCHE INTERNATIONALE

demande internationale n°

PCT/FR 99/ 00547

## Cadre I Observations - lorsqu'il a été estimé que certaines revendications ne pouvaient pas faire l'objet d'une recherche (suite du point 1 de la première feuille)

Conformément à l'article 17.2(a), certaines revendications n'ont pas fait l'objet d'une recherche pour les motifs suivants:

1. ☐ Les revendications n<sup>os</sup> se rapportent à un objet à l'égard duquel l'administration n'est pas tenue de procéder à la recherche, à savoir:
2. ☐ Les revendications n<sup>os</sup> se rapportent à des parties de la demande internationale qui ne remplissent pas suffisamment les conditions prescrites pour qu'une recherche significative puisse être effectuée, en particulier:
3. ☐ Les revendications n<sup>os</sup> sont des revendications dépendantes et ne sont pas rédigées conformément aux dispositions de la deuxième et de la troisième phrases de la règle 6.4.a).

## Cadre II Observations - lorsqu'il y a absence d'unité de l'invention (suite du point 2 de la première feuille)

L'administration chargée de la recherche internationale a trouvé plusieurs inventions dans la demande internationale, à savoir:

voir feuille supplémentaire

1. ☐ Comme toutes les taxes additionnelles ont été payées dans les délais par le déposant, le présent rapport de recherche internationale porte sur toutes les revendications pouvant faire l'objet d'une recherche.
2. ☒ Comme toutes les recherches portant sur les revendications qui s'y prêtaient ont pu être effectuées sans effort particulier justifiant une taxe additionnelle, l'administration n'a sollicité le paiement d'aucune taxe de cette nature.
3. ☐ Comme une partie seulement des taxes additionnelles demandées a été payée dans les délais par le déposant, le présent rapport de recherche internationale ne porte que sur les revendications pour lesquelles les taxes ont été payées, à savoir les revendications n<sup>os</sup>
4. ☐ Aucune taxe additionnelle demandée n'a été payée dans les délais par le déposant. En conséquence, le présent rapport de recherche internationale ne porte que sur l'invention mentionnée en premier lieu dans les revendications; elle est couverte par les revendications n<sup>os</sup>

Remarque quant à la réserve

- ☐ Les taxes additionnelles étaient accompagnées d'une réserve de la part du déposant.
- ☐ Le paiement des taxes additionnelles n'était assorti d'aucune réserve.



SUITE DES RENSEIGNEMENTS INDICUES SUR PCT/ISA/ 210

L'administration chargée de la recherche internationale a trouvé plusieurs (groupes d') inventions dans la demande internationale, à savoir:

1. Revendications: 1-47, 52-57 (complet)

Des procédés d'identification et/ou de clonage de régions d'acides nucléiques représentatives de différences qualitatives entre deux échantillons (patho)biologiques, des oligonucléotides bons pour le clonage de ces acides nucléiques, des acides nucléiques identifiés par ces procédés, des compositions et des banques comprenant ces acides nucléiques, des kits comprenant ces banques, l'utilisation de ces acides nucléiques, compositions ou banques, un procédé d'identification et/ou de production de protéines reliées à une condition pathologique, une protéine identifiée ou produite par ce procédé, un anticorps dirigé contre cette protéine, un procédé d'identification et/ou de clonage de gènes suppresseurs de tumeurs, un procédé de détermination de la toxicité d'un composé, un procédé de détermination de l'efficacité thérapeutique d'un composé, et un procédé de détermination de la réponse d'un patient à un composé.

2. Revendications: 48-51 (complet)

Une protéine de séquences SEQ ID NO:9, une sonde nucléique ou anticorps permettant d'identifier cette protéine, un procédé de criblage, et un vecteur comprenant une séquence codant pour cette protéine.

# RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux membres de familles de brevets

Document internationale No

PCT/FR 99/00547

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
WO 9527052 A	12-10-1995	AU 2237395 A CA 2186854 A EP 0775202 A JP 9511400 T ZA 9502639 A	23-10-1995 12-10-1995 28-05-1997 18-11-1997 02-01-1996
WO 9626272 A	29-08-1996	AU 5172596 A CA 2212846 A EP 0812356 A	11-09-1996 29-08-1996 17-12-1997
WO 9802576 A	22-01-1998	AU 3494297 A	09-02-1998
FR 2664287 A	10-01-1992	AUCUN	
EP 0330781 A	06-09-1989	JP 1218590 A CA 1313151 A DE 3855076 D DE 3855076 T	31-08-1989 26-01-1993 11-04-1996 17-10-1996
EP 0709397 A	01-05-1996	US 5688918 A CA 2150994 A JP 8081500 A US 5726024 A	18-11-1997 15-12-1995 26-03-1996 10-03-1998
WO 9412631 A	09-06-1994	AT 136940 T CA 2139410 A CA 2149635 A DE 69302276 D DE 69302276 T EP 0651822 A EP 0672130 A WO 9402633 A JP 2800850 B JP 8500731 T JP 8506801 T US 5830646 A US 5879898 A	15-05-1996 03-02-1994 09-06-1994 23-05-1996 19-09-1996 10-05-1995 20-09-1995 03-02-1994 21-09-1998 30-01-1996 23-07-1996 03-11-1998 09-03-1999
WO 9746679 A	11-12-1997	AU 3300397 A	05-01-1998
EP 0806478 A	12-11-1997	US 5747650 A US 5726024 A	05-05-1998 10-03-1998
WO 9630512 A	03-10-1996	FR 2732348 A AU 5402096 A BR 9607928 A CA 2214451 A CZ 9703080 A EP 0817845 A HU 9801221 A JP 11503011 T NO 974449 A SK 131197 A ZA 9602506 A	04-10-1996 16-10-1996 09-06-1998 03-10-1996 14-01-1998 14-01-1998 28-08-1998 23-03-1999 26-09-1997 06-05-1998 01-10-1996
US 5679541 A	21-10-1997	AUCUN	
WO 9704092 A	06-02-1997	FR 2736915 A	24-01-1997

# RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux membres de familles de brevets

C .nde internationale No

PCT/FR 99/00547

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
WO 9704092 A		AU 6618696 A	18-02-1997
		CA 2224468 A	06-02-1997
		CZ 9800144 A	15-04-1998
		EP 0839194 A	06-05-1998
		NO 980203 A	10-03-1998
		SK 6398 A	09-09-1998
EP 0791660 A	27-08-1997	JP 10000100 A	06-01-1998